

How to publish in *Nature*

Hopeful authors of research papers will help themselves by reading our new Guide to Authors before submission. This also highlights the need and opportunity to communicate more effectively with important non-specialist audiences.

Editors at *Nature* frequently receive passionate pleas from aspiring authors that publication of their paper is crucial for a grant, a position or for some other prospect. The scientific profession to which these authors belong is locked into a reward system based excessively on quantitative assessments, such as the number of papers published in journals that have a particular impact factor.

This broadening of the role of leading journals raises the stakes for authors. They are used to writing for fellow scientists within the discipline who will pore over every word of the methods section and have little patience with perceived hyperbole about the greater meaning of the results. Publication in these journals also requires authors to make the most of the superb opportunity to communicate to an audience far wider than their immediate peers.

Authors who submit to an interdisciplinary journal have to consider a second readership: scientists outside the discipline. These scientists read the paper for other reasons, such as a simple interest in the breadth of science, teaching, applying a new technique or observation to their own system, or ambitions to apply their discipline's armoury to the scientific challenge discussed. *Nature* has long guided its authors in this process, as their sometimes excruciatingly technical initial manuscripts travel the road towards publication in a more readable form. But authors can do more for intelligibility themselves at the outset, by showing papers to researchers with other backgrounds before submitting.

Nature provides researchers with other help, too, in the form of greatly increased efficiency and transparency in its handling of

papers, via a new web-based submission system that allows authors to track the progress of their papers through the refereeing process. We have also revised our Guide to Authors (see page 868) to provide clearer advice about how to write a paper. And we now actively encourage authors to be transparent by identifying which co-author contributed what to the work, in a section published at the end of the paper, and ensure that all authors are signed up to our principles of data access and materials sharing (see www.nature.com/nature/submit/policies for details of these and other improvements).

In addition, *Nature* invites authors to help us present their results to diverse audiences. In submitting a paper, authors are now asked for two summaries: one to summarize their work for readers (mainly scientists and editors), and another to crystallize the importance of their work for the general public. *Nature* will draw on these in the presentation and promotion of its papers.

Researchers are increasingly recognizing their duties towards broader readerships, not least the media and the more scientifically interested public. And where science can get distorted or smothered, researchers cannot sit back and let the journals do all the work for them. True, *Nature's* role is to publish the most innovative and influential papers that scientists can produce, and to present these results to the public. But it is the responsibility of researchers to seize the initiative, and communicate their knowledge and uncertainties to avoid misconceptions — and sometimes even to campaign to get science's messages across. *Nature* will continue to guide the media to authors of papers so that they are centrally involved in these opportunities. ■

In support of xeno-optimism

Despite recent gloom, there are worse things on which to spend personal wealth than a hunt for intelligent extraterrestrials.

How easy it is to scoff at 'desperate' seekers of aliens, as the science-fiction writer Brian Aldiss did in these pages two years ago (*Nature* 409, 1080; 2001). He touched some nerves: aliens are indeed as real as ghosts or numerous deities (in other words, they continue to be purely imaginary); their fictitious portrayals impede understanding; their air of being the product of scientific thinking is indeed spurious. But when all was said and done, his put-down, although erudite, was ultimately no more than a stimulating polemic.

Other prominent sceptics have been more soberingly analytical. Peter Ward and Donald Brownlee, in their book *Rare Earth* (Copernicus/Springer, 2000), discussed various factors that, on the face of it, conspire to make the emergence of life around stars probable but the appearance of intelligent species almost impossible. "Is that it?", one might ask when faced by the likelihood that we are, alas, the pinnacle of our Galaxy's intelligence. And according to a theory discussed at last week's meeting of the American Association for the Advancement of Science, if it wasn't for the chance mutation of the *FOXP2* gene 50 millennia ago, we wouldn't have the creativity to explore these ideas.

Yet more cold water has been poured on our hopes that we have intelligent company in a recent collection of answers to Enrico Fermi's pointed question about the lack of visitors: "Where is everybody?"

In a sceptical book of that title (Copernicus/Springer, 2002), Stephen Webb's explanations for the absence of alien visitors range from: "They are here and they call themselves Hungarians", attributed to Budapest-born physicist Leo Szilard, to: "Science is not inevitable", science being one of many improbable developmental steps from primitive organisms to interstellar communication or travel.

A pox on such pan-Galactic pessimism. These arguments may give sober government agencies an excuse to stop funding searches for extraterrestrial intelligence (SETI), as the US Congress did over a decade ago. But the rest of us should look more favourably on such expressions of the fundamental yearnings of humanity. All credit to the likes of William Hewlett, David Packard and Paul Allen, whose funds have allowed the SETI projects to establish new technologies — albeit still pitifully insensitive if we are to detect the equivalent of *Friends* leaking through some planetary ionosphere 1,000 light years away. All credit too to the tens of researchers who devote themselves to the dispiriting quest for such electromagnetic detritus. The rest of us should drag ourselves away from revivals of *ET*, *Close Encounters* and *Taken* long enough to scan the latest ambitions of the SETI institute, outlined in *SETI 2020* (<http://www.seti.org>), and send it a donation for the hunt for the real Thing. ■



Statement on the consideration of biodefence and biosecurity

As discussed in a Commentary by Tony Fauci on page 787, the threat of bioterrorism requires active consideration by scientists. On 9 January 2003, the US National Academy of Sciences held a discussion meeting on the balance between scientific openness and security (see *Nature* 421, 197; 2003). The next day, a group of editors met to discuss the issues with specific reference to the scientific publication process. The following statement has emerged from that meeting. The statement was conceived in a US context, but the principles discussed will be considered and followed through by *Nature* and its related journals in their international arenas.

The process of scientific publication, through which new findings are reviewed for quality and then presented to the rest of the scientific community and the public, is a vital element in our national life. New discoveries reported in research papers have helped improve the human condition in myriad ways: protecting public health, multiplying agricultural yields, fostering technological development and economic growth, and enhancing global stability and security.

But new science, as we know, may sometimes have costs as well as benefits. The prospect that weapons of mass destruction might find their way into the hands of terrorists did not suddenly appear on 11 September 2001. A policy focus on nuclear proliferation, no stranger to the physics community, has been with us for many years. But the events of 11 September brought a new understanding of the urgency of dealing with terrorism. And the subsequent harmful use of infectious agents brought a new set of issues to the life sciences. As a result, questions have been asked by the scientists themselves and by some political leaders about the possibility that new information published in research journals might give aid to those with malevolent ends.

Journals that dealt especially with microbiology, infectious agents, public health and plant and agricultural systems faced these issues earlier than some others, and have attempted to deal with them. The American Society for Microbiology, in particular, urged the National Academy of Sciences to take an active role in organizing a meeting of publishers, scientists, security experts and government officials to explore the issues and discuss what steps might be taken to resolve them. In a one-day workshop at the Academy in Washington on 9 January 2003, an open forum was held for that purpose. A day later, a group of journal editors, augmented by scientist-authors, government officials and others, held a separate meeting designed to explore possible approaches.

What follows reflects some outcomes of that preliminary discussion. Fundamental is a view, shared by nearly all, that there is information that, although we cannot now capture it with lists or definitions, presents enough risk of use by terrorists that it should not be published. How and by what processes it might be identified will continue to challenge us, because — as all present acknowledged — it is also true that open publication brings benefits not only to public health but also in efforts to combat terrorism.

The statements

First: The scientific information published in peer-reviewed research journals carries special status, and confers unique responsibilities on editors and authors. We must protect the integrity of the scientific process by publishing manuscripts of high quality, in sufficient detail to permit reproducibility. Without independent verification — a requirement for scientific progress — we can neither advance biomedical research nor provide the knowledge base for building strong biodefence systems.

Second: We recognize that the prospect of bioterrorism has raised legitimate concerns about the potential abuse of published

information, but also recognize that research in the very same fields will be critical to society in meeting the challenges of defence. We are committed to dealing responsibly and effectively with safety and security issues that may be raised by papers submitted for publication, and to increasing our capacity to identify such issues as they arise.

Third: Scientists and their journals should consider the appropriate level and design of processes to accomplish effective review of papers that raise such security issues. Journals in disciplines that have attracted numbers of such papers have already devised procedures that might be employed as models in considering process design. Some of us represent some of those journals; others among us are committed to the timely implementation of such processes, about which we will notify our readers and authors.

Fourth: We recognize that on occasions an editor may conclude that the potential harm of publication outweighs the potential societal benefits. Under such circumstances, the paper should be modified, or not be published. Scientific information is also communicated by other means: seminars, meetings, electronic posting, etc. Journals and scientific societies can play an important role in encouraging investigators to communicate results of research in ways that maximize public benefits and minimize risks of misuse. ■

JOURNAL EDITORS AND AUTHORS GROUP:

Ronald Atlas, President, American Society for Microbiology (ASM), and Editor, *CRC Critical Reviews in Microbiology*

Philip Campbell, Editor, *Nature*

Nicholas R. Cozzarelli, Editor, *Proceedings of the National Academy of Sciences*

Greg Curfman, Deputy Editor, *New England Journal of Medicine*

Lynn Enquist, Editor, *Journal of Virology*

Gerald Fink, Massachusetts Institute of Technology

Annette Flanagan, Managing Senior Editor, *Journal of the American Medical Association*, and President, Council of Science Editors

Jacqueline Fletcher, President, American Phytopathological Society

Elizabeth George, Program Manager, National Nuclear Security Administration, Department of Energy

Gordon Hammes, Editor, *Biochemistry*

David Heyman, Senior Fellow and Director of Science and Security Initiatives, Center for Strategic and International Studies

Thomas Inglesby, Editor, *Biosecurity and Bioterrorism*

Samuel Kaplan, Chair, ASM Publications Board

Donald Kennedy, Editor, *Science*

Judith Krug, Director, Office for Intellectual Freedom, American Library Association

Rachel Levinson, Assistant Director for Life Sciences, Office of Science and Technology Policy

Emilie Marcus, Editor, *Neuron*

Henry Metzger, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Stephen S. Morse, Columbia University

Alison O'Brien, Editor, *Infection and Immunity*

Andrew Onderdonk, Editor, *Journal of Clinical Microbiology*

George Poste, Chief Executive Officer, Health Technology Networks

Beatrice Renault, Editor, *Nature Medicine*

Robert Rich, Editor, *Journal of Immunology*

Ariella Rosengard, University of Pennsylvania

Steven Salzberg, The Institute for Genome Research

Mary Scanlan, Director, Publishing Operations, American Chemical Society

Thomas Shenk, President-Elect, ASM, and Past Editor, *Journal of Virology*

Herbert Tabor, Editor, *Journal of Biological Chemistry*

Harold Varmus

Eckard Wimmer, State University of New York at Stony Brook

Keith Yamamoto, Editor, *Molecular Biology of the Cell*



Ghost writers
The mystery of the paper the authors hadn't heard about p775



Goodbye, Dolly
Lung disease gets the better of the first cloned mammal p776



In the money
Donation boosts Israeli-Palestinian collaboration p777



Jump to it
US institute urged to sequence wallaby genome p778

Error reports threaten to unravel databases of mitochondrial DNA

Carina Dennis

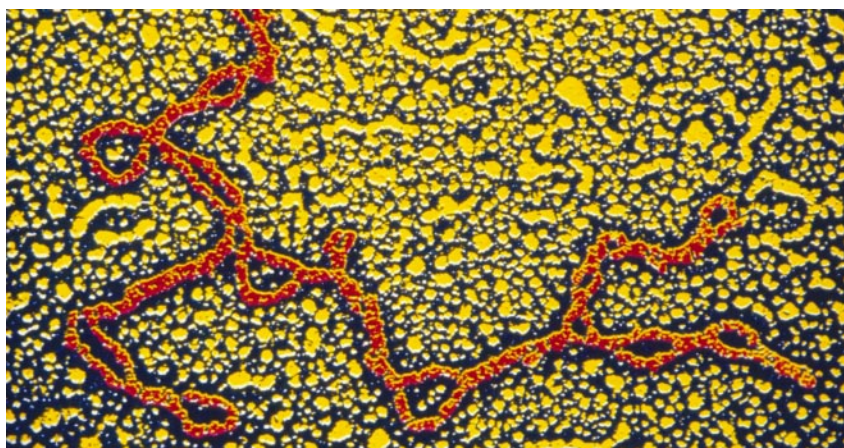
More than half of all published studies of human mitochondrial DNA (mtDNA) sequences contain mistakes, according to a geneticist at the University of Cambridge.

To the occasional chagrin of his peers, Peter Forster has repeatedly pointed out errors in published mtDNA sequences, the genetic material from cells' mitochondria, which are inherited from the mother. But his commentary in the latest issue of *Annals of Human Genetics*¹ argues that the problem is far bigger than researchers had imagined.

The mistakes may be so extensive that geneticists could be drawing incorrect conclusions in studies of human populations and evolution, says Forster. They may also confuse forensic analyses that rely on the published sequences, he adds.

"I was surprised by the number of errors," says Eric Shoubridge, a geneticist at McGill University's Montreal Neurological Institute in Canada, who investigates human diseases that result from problems with mtDNA. "What concerns me most is that these errors could be compounded in the databases."

Published mtDNA sequences are popular tools for investigating the evolution and



Human-population studies could be derailed by mistakes in recorded sequences of mitochondrial DNA.

demography of human populations. Forster has been compiling a database of corrected mitochondrial sequences published since 1981, when the complete sequence of human mtDNA — known as the 'Cambridge reference sequence' — was published². His colleagues' responses when he informs them of errors are varied. "Antagonism would be an understatement in some cases," he says. But

he adds that authors are often forthcoming in resolving discrepancies in sequences.

Neil Howell, vice-president for research at MitoKor, a San Diego-based biotech company that specializes in mitochondrial disease, says that Forster's error-detection method, which involves constructing evolutionary trees based on how sequences change, may even underestimate the extent

Nobel laureate slams misconduct smear

Quirin Schiermeier, Munich

It's like a murder without a body. An alleged case of scientific misconduct has hit the headlines in Germany and Switzerland, even though no published scientific paper has been implicated.

On 11 February, the German daily newspaper *Süddeutsche Zeitung* reported allegations of bad scientific practice brought by a postdoc who used to work in the lab of 1996 Nobel prizewinner Rolf Zinkernagel.

"The damage to our reputation is huge," says Zinkernagel, who is director of the University of Zurich's Institute of Experimental Immunology.

The allegations of the former postdoc, a 35-year-old German medical scientist, centre

on a manuscript submitted to *Science* last February. She claims that data in the paper were manipulated and that the article listed her as a co-author without her consent. The paper, a study of a cancer vaccine in mice, was withdrawn within hours of her protest.

Although Zinkernagel was not a co-author on the manuscript, the mice experiments were carried out in his lab. The study was led by Martin Bachmann, chief scientific officer of the Zurich-based vaccine company Cytos Biotechnology. Zinkernagel is a member of Cytos' scientific advisory board and has long been Bachmann's scientific mentor.

But bad blood has developed between Zinkernagel and his former postdoc, who claims that Zinkernagel was wrong not to halt

the alleged manipulation of data in the study. "There was no publication and therefore no damage to science," says Zinkernagel, who feels unfairly smeared by the affair.

The university's attempts to resolve the postdoc's complaints failed. A subsequent independent investigation, however, reported that the submitted manuscript was misleading, concluding that Bachmann violated rules of good scientific practice. But Bachmann says that inadvertent errors were immediately corrected. The media response is a personal disaster for him, he says.

The postdoc was dissatisfied with the outcome and over the next few months she asked various newspapers and journals, including *Nature*, to publish the story. ■

of the errors. Howell jointly led the team that resolved minor errors in the Cambridge reference sequence³.

The extent to which the field has been affected by mtDNA sequence errors is an open question. Forster's recent commentary was commissioned to accompany a controversial paper that re-appraises mtDNA data on European populations. The paper claims to refute earlier work that says Icelanders are less genetically heterogeneous than other European groups⁴. The extent to which erroneous sequence data affect the Icelandic issue is not resolved.

Known errors have led to a retraction of conclusions in at least one high-profile case. In 1999, a team led by scientists at the University of Cambridge seemed to have overturned dogma when they claimed to have shown mixing of maternal and paternal mtDNA by genetic recombination⁵. But their results were subsequently found to be based on mistakes in the sequence.

Some scientists question the impact that most of the sequence errors identified by Forster will have on published conclusions. "I would say that it is only a small number of cases where sequencing errors would affect the results," says Vincent Macaulay, a statistician at the University of Oxford.

Perhaps the gravest concern surrounds forensic investigations. Because large numbers of mitochondria are present in cells, they are often used to identify degraded samples from which nuclear DNA cannot be obtained. But the region of mtDNA typically used in forensics — the 'control region' — is highly variable, says geneticist Douglas Wallace of the University of California, Irvine. "People don't appreciate the fact that the control region can undergo different mutations in different cells," he says. For instance, there might be differences between mtDNA from someone's blood and from the same person's hair follicle. "This erodes the reliability of forensic assays," he says.

Most errors in published mtDNA sequences are the result of poor documentation, Forster claims. "Mistakes occur between reading the sequencing output and publishing the results," he says. Journals are as much to blame as scientists, he adds, saying that editors should be more vigilant.

Forster notes that nuclear DNA sequences in public databases are also plagued by errors, and that this may be an even bigger problem, as such mistakes are more difficult to detect.

1. Forster, P. *M. Ann. Hum. Genet.* **67**, 2–4 (2003).

2. Anderson, S. et al. *Nature* **290**, 457–465 (1981).

3. Andrews, R. M. et al. *Nature Genet.* **23**, 147 (1999).

4. Abbott, A. *Nature* **421**, 678 (2003).

5. Hagelberg, E. et al. *Proc. R. Soc. Lond. B* **266**, 485–492 (1999); erratum, *Proc. R. Soc. Lond. B* **267**, 1595 (2000).

Researchers fear the future as Congress settles 2003 budget

Hannah Hoag, Washington

Four-and-a-half months into the financial year, US research agencies finally know how much money they can spend this year.

The appropriations bill for 2003, which should have been agreed last October, was finally passed by both houses of Congress on 13 February. It provides large increases for the top two science agencies — the National Institutes of Health (NIH) and the National Science Foundation (NSF). But it has left some researchers worried that these agencies will see only slim increases in 2004.

Under the final 2003 budget, the NIH, which was set to complete a historic five-year doubling of its funding this year, will get \$26.5 billion. This is just shy of the \$27.3 billion that would have completed the doubling, and represents a 13% increase over the budget for 2002. The NSF gets \$5.3 billion, an 11% increase over the 2002 budget.

The NSF increase was welcomed by some biomedical research organizations, which have recently been calling for Congress to treat the foundation as well as it traditionally treats the NIH. "For the biomedical sciences to flourish you need the contributions that the NSF makes," says Steven Teitelbaum, president of the Federation of American Societies for Experimental Biology.

But the confirmation of the figures for 2003 has put the budget proposed by President George Bush for 2004 (see *Nature* **421**,



Roscoe Bartlett:
unhappy with
science funding.

755; 2003) under fresh scrutiny. As things now stand, the president is proposing a 2% increase for the NIH next year and 3.4% for the NSF — even though the president signed a bill last December that recommended doubling the latter's funding over five years. "The level of funding proposed in the 2004 budget is far from adequate," says Congressman Roscoe Bartlett (Republican, Maryland), a member of the House Committee on Science.

There is little good news for physicists supported by the Department of Energy's Office of Science. The \$3.3 billion budget is a meagre 2% increase over last year. "It's a lousy situation," says April Burke of Washington lobbying firm Lewis-Burke Associates, but she remains hopeful that future budgets may be more generous.

The appropriations bill passed by the Congress also allowed NASA a \$500 million increase over last year — including an additional \$50 million for its investigation into the Columbia accident — bringing its total funding to \$15.4 billion.

President Bush is expected to sign the 2003 budget into law, leaving Congress free to busy itself with his 2004 proposal.

Journals tighten up on biosecurity

Erika Check, Washington

Last week's joint statement on the handling of sensitive biological information from a group of journal editors and authors (see page 771) has received a mixed response from researchers and security experts.

The statement, which was issued at the annual meeting of the American Association for the Advancement of Science in Denver, Colorado, says that the prospect of bioterrorism raises "legitimate concerns about the potential abuse of published information". It is a response to repeated calls from US government officials to address the issue (see *Nature* **421**, 197; 2003), and commits the editors to modifying papers that they believe could represent a security threat.

Some researchers say that the statement does not take a strong enough stand in the defence of scientific freedom. "It is more equivocal and less definitive than I would like to see," says Steven Block, a biophysicist at Stanford University. But the journal editors

insist that the peer-review process would not be compromised by new security concerns.

Ron Atlas, president of the American Society for Microbiology, says that in the past year the society has modified 2 out of 14,000 submitted papers for security reasons. In one, he says, an author had written that a toxin could kill 10,000 people, but that a molecular modification to it could kill a million people. The author was asked to delete the "cookbook detail" on the modified toxin. "I don't see this as censorship — I see it as an extension of the peer-review process," Atlas says.

Others say that scientists should go much further to address security concerns about life-sciences research. David Heyman, a science and security expert at the Center for Strategic and International Studies in Washington, says that the statement is "only a step" and that scientists should make changes earlier in the research process to reduce the risk of biological research being misused.

Paper retracted as co-author admits forgery

David Adam, London

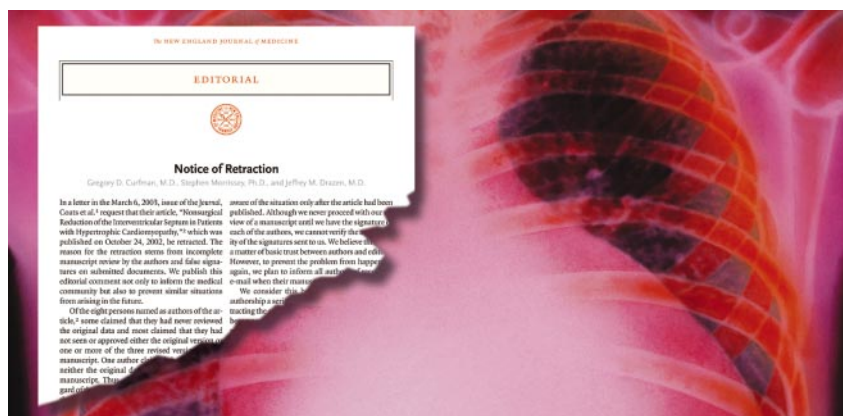
In a case more befitting Sherlock Holmes, London's Imperial College is investigating how several cardiology researchers found themselves authors of a paper that they knew little or nothing about.

The bizarre incident came to light on 10 February, when the *New England Journal of Medicine* published a retraction notice on its website. In it, the editors said that several of the eight authors on the original paper had not seen the original data or copies of the manuscript.

Suspensions were raised as soon as the paper was published on 24 October last year (W. Shamim *et al.* *New Engl. J. Med.* **347**, 1326–1333; 2002). “Over a number of days it became apparent to us that not all of the authors had been fully involved in preparing the paper,” says Gregory Curfman, executive editor of the journal.

But all eight signatures appeared on both the original submission and the three revised versions of the paper that followed. Several of these signatures, it emerged over the following weeks, were forgeries. “There were falsified signatures on the letters accompanying the original and revised versions of the manuscript,” Curfman says.

The paper describes a long-term follow-up study of 64 patients with hypertrophic cardiomyopathy, a heart condition caused by overgrowth of the muscular wall that separates the left and right ventricles. The



A paper offering hope to patients with hypertrophic cardiomyopathy is at the centre of a forgery row.

patients were treated by injecting ethanol into an artery leading to the thickened wall, which reduces the muscle's size.

Curfman puts the blame for the deception squarely on the shoulders of just one of the co-authors, who confessed in a brief letter to the journal that he had forged the signatures. But some people close to the case, including one of the co-authors who had his signature forged and did not want to be identified, point at more than one culprit.

Six of the authors on the papers were listed as having affiliations to the National Heart and Lung Institute at London's Royal Brompton and Harefield Hospital, now part of Imperial College. A spokesperson for

Imperial says that an inquiry, headed by the college's rector, Richard Sykes, is already under way. It is expected to report within three months, and the college will then decide whether disciplinary action is needed.

Nature understands that one focus of the inquiry will be the role of two of the paper's authors: Waqar Shamim, a hospital consultant described as an affiliated member of Imperial College, and Mohammed Yousufuddin, a physician who previously worked as a temporary research fellow at the Cleveland Clinic in Ohio, and who is listed as a student on Imperial's website.

Shamim is the first named author on the paper and, Curfman says, was the corresponding author on each of three previous versions. Yousufuddin is listed as the corresponding author on the published manuscript, which states that both researchers “contributed equally to this article”. Neither Shamim nor Yousufuddin, who has taken legal advice, according to some of those involved, could be traced for comment.

Hubert Seggewiss, a cardiologist at a hospital in Schweinfurt, Germany, and another of the co-authors listed on the paper, says that the article was a total surprise. “The first thing I knew of it was when Yousufuddin rang me two days before its publication to congratulate me, and to ask me about the method involved in case journalists questioned him,” Seggewiss says.

One of the pioneers of the technique described in the paper, Seggewiss says that he has performed it on over 600 patients. But he does not recognize the study in the paper and says that he has never met Shamim or Yousufuddin. “They probably used my name because in the cardiology field I am famous for this technique,” he says.

He says that some of the data “just don't add up”, but adds that Imperial's investigation will establish formally whether or not the conclusions are sound.

▶ www.nejm.org

Ministers back gene-crop advisers

Declan Butler, Paris

French government ministers have sprung to the defence of scientists who claim to have been harassed and threatened with violence after they authored a controversial report for the Academy of Sciences on transgenic plants.

In a joint statement issued earlier this month, Claudie Haigneré, junior minister for research and technology, and Luc Ferry, minister for youth, national education and research, condemned the attacks “without reserve”. They said that the methods of intimidation — at which they expressed “astonishment and sadness” — were unacceptable and were an attempt to stifle open debate on the issue of transgenic crops.

The report was submitted to the science ministry last December. It called for “reasoned and careful” introduction of transgenic crops, on a case-by-case basis, as well as an increase in research into the crops that is commensurate with their agricultural and industrial importance.

These conclusions drew criticism from opponents of transgenic crops, who challenged the report's independence. Critics accuse the authors of being active proponents of transgenic technologies. But the ministerial report says that such claims are an attempt to discredit the report's conclusions, and demonstrate “a confession of ignorance or weakness” in the arguments of those who made the threats.

Writing in French newspaper *Le Monde* earlier this year, academy president Etienne-Emile Beaulieu called on the government to defend “the honour of scientists attacked in their mission of delivering independent and educated information to society”.

Roland Douce, director of the Institute of Structural Biology in Grenoble and the report's main author, has been a principal target for the threats. He told *Nature* that such was the ferocity of the critical reaction that he would now think twice before giving public advice in the future.

Dolly's death leaves researchers woolly on clone ageing issue

Jim Giles and Jonathan Knight

The fate of Dolly the sheep is now certain: with the post mortem complete, her body will be stuffed and put on display at the Museum of Scotland in Edinburgh, UK. But her death has drawn attention to how little is known about the health problems that other clones may or may not suffer.

Dolly was put down on 14 February, suffering from a virus that caused a tumour in her lung. Ian Wilmut, leader of the team that created Dolly at the Roslin Institute near Edinburgh, says that the post mortem revealed no other gross abnormality, apart from her arthritis, which was diagnosed last year.

Dolly, born six years ago, was the first mammal to be cloned, but since then six other mammalian species — cow, mouse, pig, rabbit, goat and domestic cat — have also been cloned. The cloning process is inefficient: only around 3% of cell nuclei that are transferred to donor eggs result in live births. But no large-scale follow-ups of those births have been done, so little is known about whether clones really are likely to die young.

The health of adult clones has been studied most extensively in mice. At least 200 live mouse clones have been created in 10 or so labs around the world. In one study, Narumi Ogonuki of the National Institute of Infectious Diseases in Tokyo tracked the health of 12 mouse clones and found that 10 died — probably of pneumonia or liver disease — before their average natural lifespan of 800 days (N. Ogonuki *et al. Nature Genet.* 30, 253–254; 2002). But other researchers say that problems seen at birth or shortly afterwards, such as obesity, can reverse as the animals age. The offspring of other cloned mice have also been reported to be normal (K. L. K. Tamashiro *et al. Nature Med.* 8, 262–267; 2002).

Researchers say that experiments with larger numbers of animals are needed, but are difficult to carry out given the

difficulties associated with creating clones. Work on some animals is hampered by a lack of time and money. "You need to follow them for life. We're talking 15-plus years for cows," says Keith Campbell, a cloning researcher at the University of Nottingham, UK. ■



Get stuffed: Dolly will stay in the public eye.

Poor farmers warned against Internet transgenic crop deals

Rex Dalton

From a farmhouse in Northern Ireland, a married couple is using the Internet to tempt farmers in developing nations into planting crops that are genetically engineered to produce commercially useful molecules.

Environmentalists and scientists have expressed concern at the couple's attempts to get poor farmers from Asia to Africa to agree to make thousands of hectares available to biotechnology companies.

Critics of the plan say that it could cause an ecological disaster if local plants or crops became contaminated with transgenic strains, as developing countries often have no governmental resources for monitoring the spread of transgenes. They fear that farmers could be held responsible in the event of any such outbreak.

Brian and Diane Marshall, who live on an 80-hectare farm in Newtowncunningham near Londonderry, have used suggestions of huge monetary returns to encourage farmers from around the world to sign up to their year-old website, www.molecularfarming.com. They hope to broker contracts between the farmers and biotech firms seeking new regions to grow biopharmaceutical crops such as genetically engineered maize and tobacco. They are also seeking land in developed nations, but environmentalists are less concerned about countries that already have strict regulations for monitoring transgenic organisms.

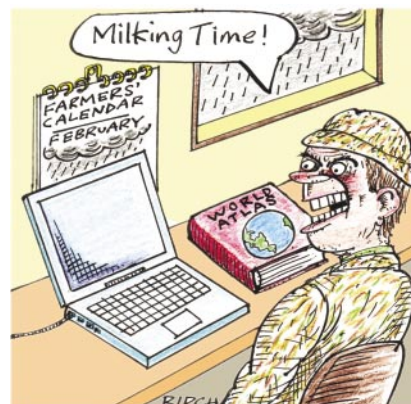
The couple, who are also seeking to collaborate with universities, insist that they have no financial arrangement with any major pharmaceutical or agricultural firm.

The Edmonds Institute, an environmental organization based near Seattle, Washington, issued an alert about the enterprise last week. "I'm concerned about the ecosystems where these crops may be introduced," says Beth Burrows, the institute's director. She fears that "small farmers in out-of-the-way places" will be held responsible if something goes wrong.

Brian Marshall retorts that the crops will not be grown for food and that potential biopharmaceutical crops will be selected to minimize the possibility of ecological damage.

But some academic scientists who are engaged in research on pharmaceutical crops are steering clear of Marshall's enterprise, fearing that the initiative could damage their research and development efforts.

"We wouldn't do anything with him," says Charles Arntzen, a plant geneticist at Arizona State University in Tempe with whom Mar-



shall had attempted to forge a collaboration. Arntzen's research involves genetically engineering plants to produce human vaccines. "This is dangerous stuff to poor people in poor countries. And it would be political death for getting research money in the future," Arntzen says.

Marshall also sought out the biotech company ProdiGene, based in College Station, Texas, as a potential collaborator. ProdiGene was fined US\$250,000 by the US Department of Agriculture (USDA) last December after its biopharmaceutical maize in Nebraska and Iowa contaminated other field crops. Anthony Laos, ProdiGene's chief executive, says that the company is not currently collaborating with Marshall, but adds: "That's not to say we wouldn't in the future."

Environmentalists in developing nations are shocked by the Marshalls' plan. Devinder Sharma, a plant geneticist and chair of the Forum for Biotechnology and Food Security in New Delhi, India, calls it "part of the global design to translocate dirty industry to the Third World".

"We cannot allow this," Sharma continues. "Let us not put the poor farmer in another trap that will land him in serious trouble."

Brian Marshall has no formal scientific training; Diane Marshall is a business professor at the North West Institute of Further and Higher Education in Londonderry. They launched their venture after changes to European regulations reduced their income from their sheep and cattle farm.

"We are full-time farmers fed up with being squeezed out of traditional farming income," Brian says. But Norman Ellstrand, a plant geneticist at the University of California, Riverside, sees the Marshalls' plan as "a frightening possibility" for the developing world's farmers. ■

Cosmologists look forward to clearer picture

Tony Reichhardt, Washington

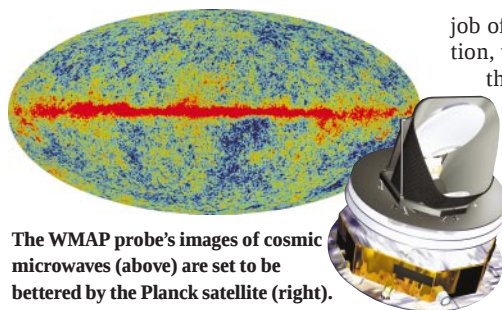
The first data from the Wilkinson Microwave Anisotropy Probe (WMAP) were released last week, some eight years after the mission was first proposed to NASA. The analysis of the data, by contrast, is happening much faster — it is already appearing on preprint servers.

The satellite, which has now spent more than a year surveying the sky at a distance of some 1.5 million kilometres from Earth, measures variations in the temperature and polarization of the cosmic microwave background (CMB), the faint radiation dating from the dawn of the Universe.

In an analysis done before the data were released, the WMAP team produced the best estimates yet of key cosmological parameters.

The new data imply an age for the Universe of 13.7 billion years, and a distribution of mass and energy in which 4% of the Universe is normal matter (atoms), 23% is dark matter, and 73% is dark energy. These figures fit well with previous estimates. The age at which the first stars seem to have formed — 200 million years after the Big Bang — is significantly earlier than most astronomers had thought.

"The really huge impact is that they tightened all the screws and bolts, and hammered everything with a superb precision," says Max Tegmark, a cosmologist at the University of Pennsylvania in Philadelphia. As a result,



The WMAP probe's images of cosmic microwaves (above) are set to be bettered by the Planck satellite (right).

he says, cosmology has taken "a giant leap forward in credibility", and "has been transformed into a real hard science".

Soon after the data were released on 11 February, papers began appearing on the arXiv preprint server operated by Cornell University. A group led by Carlos Martins of the University of Cambridge used the WMAP data to investigate the value of fundamental constants in the early Universe. Others compared the data to previous sky surveys and assessed its impact on theories of the geometry of the Universe. "I predict that you'll see a thousand other papers in the coming years," says Tegmark.

The next big step in all-sky CMB mapping will be the European Space Agency's Planck satellite, due to launch in 2007. Planck will have three times the resolution of WMAP, and it should do a more thorough

job of characterizing the foreground radiation, which can mask the faint signal from the microwave background.

Before then, several ground-based studies will examine small patches of the sky with even greater resolution. Among these are the South Pole Telescope, led by the University of Chicago, and the Chilean-based Atacama Pathfinder Experiment, proposed by the Max Planck Institute for Radioastronomy in Bonn. Their images of the microwave background may yield new information on the period when the first stars formed.

Another goal for future investigations is a more precise measurement of the polarization of the microwave background, discovered last year by the Degree Angular Scale Interferometer in Antarctica. The WMAP also measured this polarization, and a host of ground- and balloon-based instruments will take the investigation a step further.

One such balloon-based experiment, called BOOMERANG, recently completed a flight in Antarctica, where it collected "beautiful data" on the polarization with greater sensitivity than WMAP, according to principal investigator Andrew Lange of the California Institute of Technology. He expects to publish the results later this year.

♦ <http://arxiv.org>

Windfall spurs rare Israeli–Palestinian research effort

David Adam, London

Most collaborations between Israeli and Palestinian scientists have fallen victim to the upsurge of violence that has disrupted the region in recent years. But one surviving partnership — an investigation of the hereditary basis of deafness among people in the region — last month received a welcome boost: a donation of US\$100,000 from a \$1-million prize scooped by geneticists John Sulston, Sydney Brenner and Robert Waterston.

The research is led by geneticists Karen Avraham, an Israeli at Tel Aviv University, and Moien Kana'an, a Palestinian who works at Bethlehem University on the West Bank. Kana'an says that they have already recruited some 75 Palestinian families for the project, which is aimed at mapping the genes that cause their society to have an unusually high rate of genetic deafness. Avraham hopes that Sulston's donation will allow them to bring in new postgraduate students in October to join Hashem Shahin, a Palestinian PhD student who has worked in her lab since 1998.

The donation comes from the winnings collected by Sulston, Brenner and Waterston

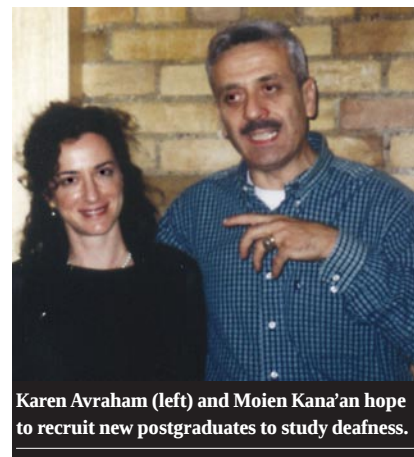
after being awarded one of the three Dan David Prizes in 2002. These prizes, set up by Dan David, president of photo-booth operators Photo-Me International, recognize outstanding "scientific, technological, cultural or social impact on our world". The award's winners are compelled to give 10% of their prize to younger colleagues in their field.

"We've wanted to expand the scheme for years," Avraham says. It was the political situation rather than a lack of funds that was putting the duo off, she adds, but Sulston's move has persuaded them to take the plunge. "It's a validation of what we're doing and has given us hope," she says.

Both Avraham and Kana'an stress that their cooperation is for the benefit of science. "The collaboration itself was not a political decision," Kana'an says. "It started at a scientific grassroots level and has developed from there." Kana'an is the only Palestinian researcher investigating the problem locally, although several international groups have visited the region to take DNA samples.

One feature of the study is that many of the families recruited are very large and

have more than one affected member, making it easier to track common genes that may be responsible. "Very large families are incredibly useful in this type of research," says Karen Steel, who studies hereditary deafness at the Medical Research Council's Institute of Hearing Research in Nottingham, UK. "They are a very powerful resource for gene mapping."



Karen Avraham (left) and Moien Kana'an hope to recruit new postgraduates to study deafness.

Epidemiologists warn of widespread arsenic problem in India

New Delhi The arsenic contamination of well water on the Indian subcontinent could affect much greater numbers of people than had previously been thought, a study suggests.

Arsenic was first spotted in well water in the 1980s on the Bengal Delta, an area shared by India and Bangladesh that is the coastal flood plain of several rivers including the Ganges. But Dipankar Chakraborti, an epidemiologist at Jadavpur University in Kolkata, India, has now confirmed reports of arsenic poisoning in the Indian state of Bihar, 500 kilometres west of the delta (D. Chakraborti *et al. Environmental Health Perspectives* doi:10.1289/ehp.5966; 2003).

Chakraborti realized that there might be a problem after hearing of a spate of cancer deaths and skin lesions in the area, both of which are associated with the chronic intake of arsenic. The study raises fears that arsenic-laden aquifers could cover much of the Ganges basin, which stretches beneath the foothills of the Himalayas from New Delhi to the Bay of Bengal.



Health problems: well water contaminated with arsenic can cause skin lesions and cancer.

Göttingen rapped by German funding agency

Munich The University of Göttingen has been accused of ignoring a request to set up an independent inquiry into allegations of scientific misconduct.

A letter to the university from the DFG, Germany's main grant-giving agency and ethical guardian for universities, was leaked last week to the *Süddeutsche Zeitung* newspaper. In the letter, the DFG complains that the university has ignored a request from the agency's ombudsman to set up an investigation into a paper on the skin disorder neurodermatitis. The study by V. Schettler *et al. (Kidney and Blood Pressure Research* 24, 213–440; 2001) claimed clinical success for a new immunological treatment for the disorder.

The university's medical faculty investigated the study last year after questions were raised about the data involved. But the DFG said that the faculty's conclusion — that misconduct could not be ruled out and that a retraction was “expected” — was too mild.

The University of Göttingen has recently been criticized for its slow handling of a high-profile study of a cancer vaccine published in *Nature Medicine* (see *Nature* 420, 258; 2002).

Japan matches price of German collider

Tokyo Japanese physicists have unveiled the most detailed cost estimates yet of their version of a next-generation electron–positron collider.

The Japanese Electron–Positron Linear Collider Project calls for the construction of a 30-kilometre-long pair of linear accelerators to smash electrons and positrons together at energies of 500 GeV. Depending on the version of the collider to be built, its cost will be between US\$2.5 billion and US\$4.2 billion, Japanese researchers announced on 12 February.

This puts the project's cost in line with a rival German proposal, known as TESLA, estimated at US\$3.7 billion. The high-energy collisions are needed to illuminate the behaviour of particles and forces at extremely high energies and over very small distance scales.

♦ <http://www.jlc.kek.jp>

♦ <http://tesla.desy.de>

Europe uproots patent for DuPont's oil-rich maize

Munich A patent on a variety of maize with a high yield of oil has been revoked by the European Patent Office, following challenges by the Mexican government, Greenpeace and Misereor, an overseas-development agency run by the German Catholic Church.

The maize was developed by the chemical and agricultural company DuPont, based in Wilmington, Delaware. Pollen was exposed to a chemical mutagen and then cross-pollinated with normal maize. The resulting plants were then screened for high oil yields — the patent was awarded to a variety with an unusually high proportion of oleic acid, the predominant oil in olive oil.

Opponents argued that the patent was not valid because DuPont's techniques are widely used, and because maize with high yields of oil have been developed through normal plant-breeding programmes. The Mexican government also objected, although the patent would have had no legal grounds in the country. A spokesperson for the patent office says that farmers would still have been able to use existing high-yield crops.

Gravitational waves remain out of reach

Denver An ambitious US project to detect gravitational waves presented its first results this week at the American Association for the Advancement of Science's annual meeting in Denver, Colorado. The data were collected over 17 days in September by the two facilities of the Laser Interferometer Gravitational-Wave Observatory. No waves were seen, but the measurements will help astronomers predict the power and frequency of the high-energy events that are thought to create detectable gravitational waves.

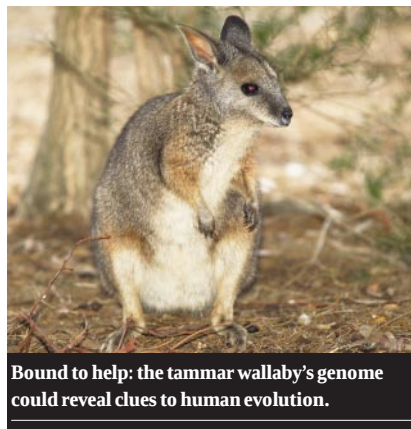
The observatories, based in Livingston, Louisiana, and Hanford, Washington, have been trying to reduce noise picked up by their detectors (see *Nature* 417, 482–484; 2002), and it is unclear whether the current designs will be able to detect gravitational waves. Plans for an upgrade costing US\$100 million to US\$150 million were unveiled this week.

Wallaby hops into running for genome sequence

Sydney A rabbit-sized kangaroo called the tammar wallaby (*Macropus eugenii*) is to enter the competition for time and money at US gene-sequencing centres. If successful, it will join the chimpanzee and honeybee in the queue of organisms waiting to have their genomes read letter by letter.

Geneticist Jenny Graves of the Australian National University in Canberra, who is spearheading the kangaroo campaign, says that in terms of our evolution the kangaroo lies between close cousins such as the mouse and more distant relatives such as birds and fish. As such, it could generate new insights into how the human genome evolved. Graves and others have used kangaroo genetics to show how human sex chromosomes evolved from standard chromosomes, for example.

The bid was dispatched in time to meet the 10 February deadline set by the US National Human Genome Research Institute. But competition is stiff: the cat, pig and rhesus monkey are all expected to be in the running.



Bound to help: the tammar wallaby's genome could reveal clues to human evolution.

All together now



From wobbly bridges to new speech-recognition systems, the concept of synchrony seems to pervade our world. Steve Nadis reports on attempts to understand it, and the applications that may be on the horizon.

P. JORDAN/PA

CLEMSON UNIV.

Steven Strogatz's curriculum vitae is more eclectic than most. He has investigated how crickets come to chirp in harmony, and why applauding audiences spontaneously clap in unison. The theme behind such studies — the way in which systems of multiple units achieve synchrony — is so common that it has kept him busy for over two decades. "Synchrony," says Strogatz, a mathematician at Cornell University in Ithaca, New York, "is one of the most pervasive phenomena in the Universe."

When a mysterious wobble forced engineers to close London's Millennium Bridge shortly after it opened in 2000, for example, an unforeseen synchronizing effect was responsible: walkers were responding to slight movements in the bridge and inadvertently adjusting their strides so that they marched in time. But synchrony can provide benefits too: researchers working on new radio transmitters and drug-delivery systems are harnessing the phenomenon to impressive effect. "It occurs on subatomic to cosmic scales and at frequencies that range from billions of oscillations per second to one cycle in a million years," says Strogatz. "It's a way of looking at the world that reveals some amazing similarities."

The study of synchronous systems cuts across the disciplines of modern science. But the underlying phenomenon was first documented over three centuries ago. In 1665, Dutch physicist Christiaan Huygens lay ill in

bed, watching the motions of two pendulum clocks he had built. To his surprise, he detected an "odd kind of sympathy" between the clocks: regardless of their initial state, the two pendulums soon adopted the same rhythm, one moving left as the other swung right.

Elated, Huygens announced his finding at a special session of the Royal Society of London, attributing this synchrony to tiny forces transmitted between the clocks by the wooden beam from which they were suspended. But rather than inspiring his peers to seek other examples of self-synchrony, his study was largely ignored. The heir to Huygens' idea was not a seventeenth-century scientist, but Arthur Winfree, a theoretical biologist who began in the 1960s to study coupled oscillators¹ — groups of interacting units whose individual behaviours are confined to repetitive cycles.

Jungle rhythms

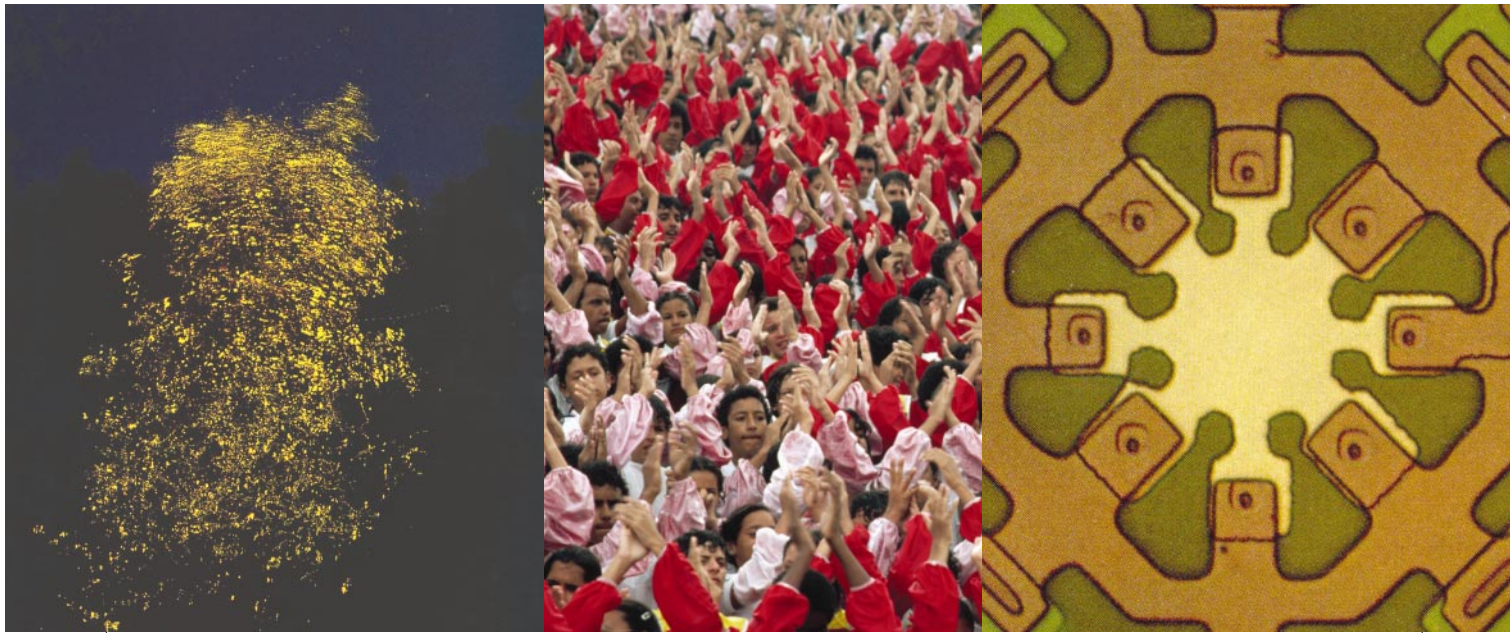
The blinking of fireflies is one behaviour that Winfree studied. As night falls on the jungles of Southeast Asia, fireflies begin to flicker, each following its own rhythm. But over the next hour or so, pockets of synchrony emerge and grow. Thousands of fireflies clustered around individual trees eventually flash as one, switching on and off every second or two to create a stunning entomological light show.

How does such synchrony come about? In this case, each firefly has its own cycle of

flashes, but that rhythm can be reset when the fly sees a flash from a neighbour. Pairs of flies become synchronized in this way, and the effect gradually spreads until large groups are linked. In general, oscillating units communicate by exchanging signals that prompt other units to alter their timing. Synchronization occurs if these 'coupling' signals are influential enough to overcome the initial variation in individual frequencies. "Below a threshold, anarchy prevails; above it, there is a collective rhythm," Winfree wrote in a review article published shortly after his death in November 2002 (ref. 2).

Winfree's attempts to create a detailed mathematical model of coupled oscillators were stymied by the difficulty of solving nonlinear differential equations — the mathematical tools used to describe such systems. But a crucial breakthrough came in 1975, when Yoshiki Kuramoto, a physicist at the University of Kyoto in Japan, produced a simplified model of the kind of system that Winfree was interested in. Kuramoto's system, in which the oscillators are nearly identical and are joined by weak links to all of the others, can be described by a set of largely solvable equations³.

Kuramoto did not assume that his abstract model would necessarily relate to real physical systems. But that changed in 1996 when Strogatz, together with physicists Kurt Wiesenfeld of the Georgia Institute of Technology in Atlanta and Pere Colet, then at



Cycling club: synchronizing systems in both natural and technological settings. Left to right: pedestrians make London's Millennium Bridge wobble; crickets and fireflies synchronize their chirps and flashes; an audience claps in sync; and the electric currents through Josephson junctions oscillate as one.

the Institute of Material Structures in Madrid, produced a mathematical description of an array of superconducting devices called Josephson junctions⁴. These consist of an insulating layer, so thin that electrical current can actually cross it, sandwiched between two superconducting metals. Once the current across the junction exceeds a certain level, the direction of flow oscillates very rapidly, sometimes exceeding 100 billion cycles per second.

According to Wiesenfeld and his colleagues, an array of junctions will come to oscillate in sync as connections between the junctions nudge the devices into phase. Electrical engineers, who hoped that Josephson junctions could be used to drive a new breed of faster computers, were intrigued by the idea. What's more, in the same paper, the trio also showed that their theoretical description is equivalent, in mathematical terms, to Kuramoto's model. The finding kick-started interest in synchronized systems, capturing the attention of researchers from across the scientific spectrum.

John Hopfield, a theoretical physicist at Princeton University in New Jersey who pioneered studies of artificial neural networks, is one example. Computer simulations of networks of simplified model neurons are known to be well suited to certain tasks, such as pattern and face recognition. But Hopfield is now working with both real and simulated networks of units that behave more like actual neurons. Each neuron in his network emits voltage pulses at regular intervals, which are relayed to other parts of the network. Like the fireflies, a neuron's firing cycle can be reset by an incoming signal, allowing groups of neurons to synchronize their outputs.

In 2001, Hopfield described how this synchrony could be exploited to create a speech-recognition device⁵. He simulated a network of 650 biologically realistic neurons with only weak couplings between them, initially using conventional sound-analysis software to divide spoken words into 40 'channels'. Each channel corresponds to a particular range of sound frequencies and one of three key events: the time at which the sound of that frequency began, when it peaked, and when it stopped. Each thus has a time associated with it, which states when a particular frequency turned on, off or peaked. Neurons in Hopfield's network are connected to one or more of these channels, firing off a series of regular pulses when they receive the time signal. The frequency of this firing decreases with time, and although this rate varies between neurons, all eventually fall silent.

One to think about

So how does such a set-up recognize sounds? Neurons are activated at different times, but because their firing frequencies fall off at different rates, some of them will momentarily fall into sync with each other before drifting out of phase again. In a first trial run, Hopfield fed the word 'one' into the network and tracked the firing of the neurons until he spotted a group that moved into phase. He then strengthened the coupling between these neurons. When the word 'one' was presented a second time, this coupling was sufficient to prompt a burst of synchronous and easily detectable firing when the neurons drifted into phase. Other words did not cause this subset of neurons to come into phase, and hence did not prompt synchronous firing.

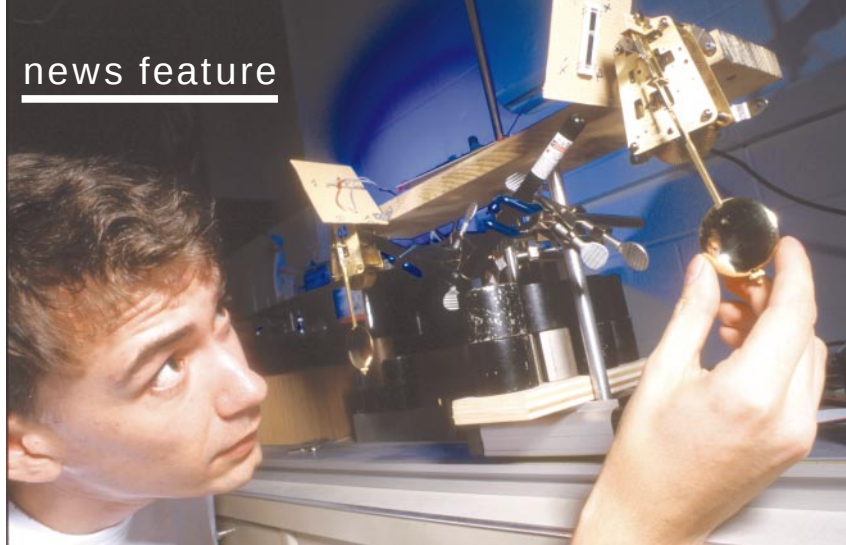
The network could speed up speech recognition, as detecting synchronous firing is much quicker than identifying a word by analysing each channel. "If you take a system that can spontaneously synchronize, you immediately get an answer: it's in sync or it's not," says Hopfield. He suggests that the approach could be useful for answering questions in tasks such as face recognition, "where you have lots of information coming in and all you really want to know is yes or no".

At the University of Pennsylvania in Philadelphia, bioengineer Kwabena Boahen has created real systems, each consisting of a network of thousands of circuits that mimic the behaviour of neurons. Theoretical studies of these networks suggest that their synchronous firing could be put to good use⁶. Boahen's circuits can be trained to recognize a particular pattern of inputs. By measuring the proportion of neurons that fire in sync, an observer can judge the degree of certainty associated with the decision. An input that causes 90% of neurons to fire in sync, for example, is more likely to have been recognized than one that causes 80% to synchronize. "This shows you can answer more than just yes/no questions," Hopfield comments. "Instead, you can ask what is the degree of confidence that this face belongs to 'Joe'?"

While Hopfield and Boahen are pursuing computational methods inspired by neural circuits, other investigators hope to exploit synchrony at the level of genes and proteins. Nancy Kopell, Jim Collins and their colleagues at Boston University in Massachusetts are trying to construct a synthetic regulatory network in the bacterium *Escherichia coli* that turns genes on and off on a periodic basis. Last year, they described a theoretical

ABOVE LEFT, S. MAZE/CORBIS; ABOVE, H. VAN DER ZANT

I. POLUNIN



Swinging time: a Georgia Tech researcher recreates Christiaan Huygens' twin pendulum experiment.

cell⁷ that contains genes for two proteins, X and Y. X activates the genes that encode both itself and Y, and this positive feedback causes levels of X and Y to rise. But in the Kopell–Collins model, Y also degrades X, so that levels of X fall as Y builds up. This in turn reduces the activity of the gene for Y. With less Y around, X levels increase and the cycle repeats itself.

Each oscillating set of genes can be coupled by introducing a third protein, A, which diffuses between cells. The gene for A is activated by X, and A in turn activates X, so levels of A and X rise together. As these levels increase, molecules of A diffuse from the cell and boost levels of X in neighbouring cells. This resets the cycle of fluctuating X levels in neighbouring cells, bringing them into line with the cell from which A originally diffused.

Theoretical analysis of a population of 1,000 cells based on biologically plausible rates of diffusion suggests that they will all fall into synchronization within a matter of minutes, even when the simulation begins with cells distributed at random points in their cycle. In experiments set to begin later this year, the Boston University team will find out whether this idea holds up in the lab. If it does, the levels of one of the proteins produced by the cell will peak around once an hour, although this frequency could be adjusted. In the long term, they hope to use a similar strategy to produce therapeutic substances at regular intervals, to form part of a drug-delivery system for use inside the body.

Evidence that this approach could work in practice comes from a 2000 paper by theoretical physicists Michael Elowitz and Stanislas Leibler, then both at Princeton University. Elowitz and Leibler created an oscillating three-gene network in *E. coli*⁸, in which the protein produced by the first gene suppresses the activity of the second gene; the second protein suppresses the third gene; and the third protein suppresses the first. In this way, levels of the three proteins successively rise and fall over a period of two to three hours. Collins and Kopell hope to build on this achievement, establishing oscillations such

as this in many cells and then getting the oscillations to synchronize.

Other examples of research into self-synchronizing systems abound. Neuroscientists are debating how synchronous neural activity within the brain influences attention, and perhaps even consciousness. Studies of the breakdown of synchronous beating among heart-muscle cells could lead to a better understanding of cardiac arrhythmias. And in 2001, Wiesenfeld and his Georgia Tech colleagues repeated Huygens' experiment under more rigorous conditions⁹, tracking the pendulums' movements with lasers, as a means of generating data for Wiesenfeld's theoretical studies of synchrony.

Wider view

Meanwhile, Strogatz is interested in expanding the range of systems that are studied under the banner of synchrony. "We've gone far by limiting our focus to repetitive behaviour," says Strogatz, whose new book on synchronization will be published next month¹⁰. But the time is ripe to loosen the shackles of the Kuramoto model, he suggests, and entertain more general conditions.

The biological circuits studied by Kopell and Collins are one example, as the signalling between the cells is stronger than the coupling that Kuramoto built into his model. Work by Robert York, an electrical engineer at the University of California, Santa Barbara, represents another step away from sim-



Steve Strogatz: it's time to study more systems.

plified oscillator networks. York has constructed a string of ten radio transmitters¹¹ — the frequency of radio waves that each emits is determined by the oscillating current that is fed into it. The circuits that produce these currents are linked, and fall into sync with each other less than a nanosecond after they are turned on.

In York's system, each transmitter is coupled only

to its nearest neighbour. But this doesn't prevent the array from synchronizing. What's more, it also allows York to control the frequency at which the array synchronized, simply by adjusting the oscillator circuits for the antennae at each end of the array.

A group headed by Brian Meadows, a physicist at the US Navy's Space and Naval Warfare Systems Command in San Diego, is scaling up this idea, preparing to build a square array of 900 radio antennae to see whether the same approach works in two dimensions. Such systems are attractive, as they are more flexible than a single large antenna and can be packed more tightly than a conventional array. If Meadows' array works, it could yield a wide variety of applications, such as compact system for ships, airliners and satellites. "Normally you can't put antennae too close, because coupling becomes a problem," says Meadows. "For us, this coupling is essential and we take full advantage of it."

But the biggest challenge may be understanding systems containing oscillators that are far from identical. "In physics, we're used to dealing with things like electrons and water molecules that are all the same," says Strogatz. "But no one knows how to deal mathematically with the tremendous diversity that biology presents." He wants to replace idealized oscillators with real biological elements such as genes and cells, but considers the task daunting. "Biologists are used to collecting as many details as possible," he says. "For someone like me, the trick is to see which details we really need. But there's no guarantee that simplification will work in our efforts to model cellular processes."

Strogatz is nevertheless convinced that such studies will one day bear fruit. "Virtually all of the major unsolved problems in science today concern complex, self-organizing systems, where vast numbers of components interact simultaneously, with each shift in one agent influencing the other," he says. Huygens had a similarly strong conviction that he had stumbled into something big, which was sufficient to rouse him from his sickbed, even if he could not have fathomed its full significance at the time. Only now are we getting a glimpse of how enduring his legacy may be. ■

Steve Nadis is a freelance writer in Boston.

1. Winfree, A. T. *J. Theor. Biol.* **16**, 15–42 (1967).
2. Winfree, A. T. *Science* **298**, 2336–2337 (2002).
3. Kuramoto, Y. *Int. Symp. Math. Problems in Theor. Phys.* (ed. Araki, H.) 420–422 (Springer, Heidelberg, 1975).
4. Wiesenfeld, K., Colet, P. & Strogatz, S. H. *Phys. Rev. Lett.* **76**, 404–407 (1996).
5. Hopfield, J. J. & Brody, C. D. *Proc. Natl Acad. Sci. USA* **98**, 1282–1287 (2001).
6. Hynna, K. & Boahen, K. *Neural Networks* **14**, 645–656 (2001).
7. McMillen, D., Kopell, N., Hasty, J. & Collins, J. J. *Proc. Natl Acad. Sci. USA* **99**, 679–684 (2002).
8. Elowitz, M. B. & Leibler, S. *Nature* **403**, 335–338 (2000).
9. Bennett, M., Schatz, M. F., Rockwood, H. & Wiesenfeld, K. *Proc. R. Soc. Lond. A* **458**, 563–579 (2002).
10. Strogatz, S. H. *Syn: The Emerging Science of Spontaneous Order* (Hyperion, New York, in the press).
11. Liao, P. & York, R. A. in *IEEE MTT-S International Microwave Symposium Digest 1235–1238* (Inst. Elec. Electron. Engin., San Diego, 1994).

To the heart of glass

How do you run an experiment that takes thousands of years? As scientists studying nuclear-waste disposal are finding out, the answer is to look back in time. Philip Ball reports.



The ravages of time on glass can be assessed by examining ancient artefacts, such as this Roman vase dating from AD 250.

If you feel that your experiments can take for ever, spare a thought for the scientists studying the disposal of nuclear waste. The half-lives of some radioactive wastes stretch to hundreds of thousands of years. How can the materials used to contain these wastes ever be properly tested?

The solution, or at least part of it, is to collaborate with archaeologists. One of the best disposal options is to trap the waste in glass and bury it. So when the civilizations of the Middle East first learnt how to make glass at least 4,500 years ago, they unwittingly launched an experiment on the long-term stability of the favoured material for storing nuclear waste. Archaeologists have since excavated samples of these glasses, and their findings could help materials scientists improve nuclear-waste disposal.

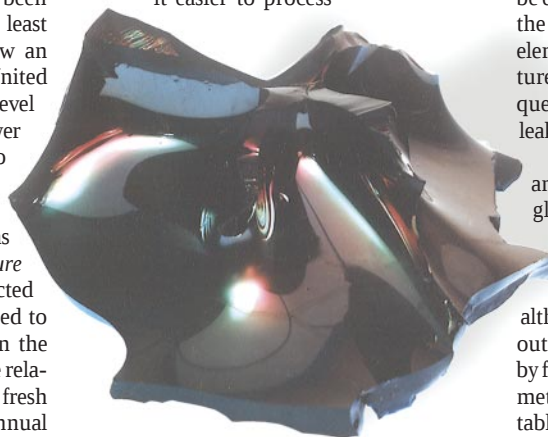
The two communities have actually been courting each other sporadically for at least two decades, but nuclear waste is now an increasingly pressing problem. In the United States alone, 400 million litres of high-level waste — spent fuel from nuclear power plants and the military reactors used to make weapons-grade plutonium — need to be stored. A proposed repository inside Yucca Mountain in Nevada has attracted much controversy (see *Nature* **412**, 850–852; 2001), although it is expected to open early next decade. With the need to test potential storage materials high on the agenda of nuclear-waste researchers, the relationship with archaeologists got off to a fresh start in Boston last December, at the annual meeting of the US Materials Research Society.

Glass is already an established part of the waste-disposal process. Since 1996, a plant at

the Savannah River Site, a former nuclear-weapons production facility in South Carolina, has been mixing high-level waste into molten glass and casting it into cylinders about a cubic metre in volume. At Yucca, these would be placed in canisters made of a corrosion-resistant nickel alloy. But water is likely to seep into the underground chambers holding the waste. No one knows for sure if the alloy will hold out over thousands of years, so materials scientists are interested in how water will affect the encased glass over long periods of time.

Waste cocktail

Normal glass consists mainly of a disorderly network of silicon and oxygen atoms. Typically incorporated into this network are small amounts of metals such as sodium and calcium, which are introduced to make it easier to process



Glass is stable, but is it sufficiently robust to store nuclear waste for hundreds of thousands of years?

the glass. Larger metallic elements, including heavy radioactive elements such as caesium and neptunium that are found in nuclear waste, can be caught and held within this silicate network. Stable and robust, glass seems to be an ideal option for storing waste.

But over the thousands of years that nuclear waste remains active, glass can undergo significant changes. For relatively short periods of time, a few years say, water barely perturbs silicate glass. But hydrogen ions in water will eventually swap places with the sodium and calcium in the glass, slowly disrupting the network and allowing some of the radioactive elements to escape.

What's more, silicate glass can transform into a crystalline form over hundreds of years, squeezing the radioactive atoms out of the crystal lattice in the process. The glass will also be exposed to high levels of radiation and, for the first few decades, heat from the radioactive elements will create defects in the glass structure that can initiate further damage. The key question is not if nuclear waste will eventually leak out, but when and how fast.

Vital clues could come from studies of ancient glass. The earliest-known artificial glass comes from Mesopotamia, now part of Syria and Iraq, and dates from 2500 BC. The ancient Egyptians also developed considerable expertise in glassmaking, although it remained a luxury item throughout the Middle Ages. These glasses were made by fusing sand with substances that contained metals from groups I and II of the periodic table, such as sodium, found in wood ash.

This is fortunate, as several of the recipes being studied for nuclear-waste glasses also contain high amounts of such metals. Some of



This disc shows how much high-level waste is produced if one person's lifetime electricity needs come from nuclear power.

the US waste is currently stored at a former nuclear-weapons facility in Hanford, Washington. Two of the glasses being tested there for low-level waste are HAN-28F and LAWA44, which contain 20–30% sodium oxide.

Moreover, metals such as cobalt and manganese that were added to ancient glass to give it different colours can serve as markers for evaluating how radionuclides in a waste glass might disperse into the surrounding earth. The biggest long-term environmental impact from leaky nuclear waste will come from just a few long-lived and soluble radionuclides, such as technetium-99 and iodine-129. The first of these, for example, has a half-life of about 200,000 years.

But archaeological studies do not provide a perfect guide to the waste-disposal glasses. Unlike ancient glasses, the latter typically contain a relatively large amount of boron or phosphorus. Moreover, the glasses used to store high-level waste do not have the high levels of group I and II metals that archaeological glasses do. "You'll never get a perfect analogue out of an ancient glass," says Pam Vandiver, a materials scientist working on the conservation of ancient artefacts at the Smithsonian Center for Materials Research and Education in Suitland, Maryland.

Despite these drawbacks, materials scientists can still make use of archaeological studies. Ancient glasses provide useful

information on how glass corrodes, even if it is not specific to waste-disposal glasses. Experience from archaeological sites suggests that there are many ways in which corrosion occurs, depending on factors such as the composition of the glass, and the acidity and temperature of the soil. Not all degradation is a result of chemical weathering: microbes can also attack the glass.

Window from the past

More importantly, archaeological studies can help researchers to improve their theoretical models and computer simulations of glass corrosion. By adapting these models to the composition of particular archaeological specimens and the environment that they were buried in, researchers can run their models and compare the predictions to real measurements taken on the ancient samples. "A similar mechanism seems to operate for all silicate glasses," says Peter McGrail, a materials scientist at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington. "So the ancient glasses provide a unique validation opportunity that is applicable to all silicate waste glasses."

It is too early to say what the results of this opportunity will show, but nuclear-waste researchers are excited about the possibilities. "Nothing beats a completely independent experiment, and that's what the ancient analogues provide," says Denis Strachan, who studies nuclear-waste materials at PNNL. "If we can verify our models that way, we can show that they are trustworthy."

Further input into the models comes from experiments in which materials scientists bury their samples and dig them up years later. Such tests are being conducted in the United States, Russia, Sweden and

Belgium. One of the most exciting field studies for nuclear-waste researchers is one that began partly with archaeological objectives in mind. In 1970, Walter Fletcher of the British Glass Industry Research Association, now part of the Sheffield-based British Glass Manufacturers' Confederation, arranged for a wide variety of different types of glass, including synthetic analogues of Roman and medieval glass, to be buried in limestone soils in a quarry in Ballidon in Derbyshire. Various modern glasses were buried too, to act as reference samples against which the simulated archaeological glasses could be compared. Fletcher intended that some of these samples would be exhumed at set intervals for centuries to come.

The Ballidon project, now overseen by researchers at the University of Sheffield, came to the attention of George Wicks, a materials scientist at the Savannah River Site, in 1985. The following year saw the fifth exhumation, and Wicks took the opportunity to bury some nuclear-waste glasses. These were exhumed last year, and the results so far are encouraging. "We found out that the corrosion of the glass is very low," says Wicks, "and the glass is doing better in the field than in lab leaching tests." The amount of leaching from the samples actually seems to decrease with time.

When he completed the final exhumation, Wicks also buried new samples, including HAN-28F and LAWA44 glasses, which are scheduled to be exhumed in 2, 5, 10 and 16 years' time. McGrail would like to see soil samples collected from the site at the same time.

It was straightforward for Wicks to include his samples in the Ballidon project, but getting materials scientists and archaeologists to work together is not always so easy. McGrail points out that although tracking trace elements is important in studies of nuclear-waste disposal, archaeologists rarely bother to check what might have leaked from their artefacts and entered the soil.

Nuclear-waste researchers, on the other hand, can use and abuse their samples with relative profligacy, but investigations of archaeological glass objects must often use non-destructive methods. Such issues serve as a reminder that, although both communities might be interested in questions such as how glass weathers, they may not always be asking the questions in the same way. The approach and methods of the two communities are almost "mutually exclusive," Vandiver admits.

Despite these differences, materials scientists are in no doubt about the value of the information that archaeological studies can provide. Without such work, they would be left with no long-term data with which to evaluate their models. No one can see a hundred thousand years into the future, but confidence in nuclear-waste glasses would receive a major boost if the predictions work out for the glass of the Pharaohs.

Philip Ball is a consultant editor of *Nature*.



Digging that scene: George Wicks (left) and his colleague test the resilience of glass in the ground.

R. NEWTON

Scientific freedom: new strategies are needed

Threats to the flow of knowledge may come from small groups as well as governments.

Sir — The International Council for Science (ICSU) welcomes your timely publication of the Commentary by Blakemore *et al.*, “Is a scientific boycott ever justified?” (*Nature* **421**, 314; 2003). We are pleased that the authors draw a conclusion directly aligned with ICSU’s ‘universality of science’ principle (see www.icsu.org/Library/Central/Statem/frecondsc.html) — that the unrestricted flow of scientific ideas and information is critical to the advancement of science — in stating that there would be few, if any, circumstances in which the principle should be overridden.

Since its inception in 1931, ICSU has affirmed the fundamental rights of scientists to pursue research and publish results, to associate freely, and to share materials. We have publicly upheld this principle through the political and social turmoil of the Second World War, the Cold War and apartheid. Our work to protect the rights of individual scientists — often undertaken out of the limelight — has been steady and successful.

As Blakemore *et al.* suggest, each situation is unique and calls for a considered response, not just from ICSU, but also from its members (the 101 national members and 27 scientific unions for which adherence to the principle is a condition of membership) and from the entire scientific community.

The Commentary was prompted by a call — made by Blakemore, his co-author Richard Dawkins and 123 other academics, in a letter to Britain’s *Guardian* newspaper on 6 April 2002 — for national and European agencies to suspend Israel’s eligibility for academic funding until Israel agreed to abide by UN resolutions and open serious peace negotiations with the Palestinians. Others subsequently called for boycotts of Israeli scientists, and two journals decided to exclude Israeli scholars from their editorial boards. (See also Correspondence by M. Fainzilber, *Nature* **417**, 15; 2002; and by S. Rose & H. Rose, A. Abbes *et al.*, and M. Mangel, *Nature* **417**, 221–222; 2002.)

These events prompted ICSU to issue a public statement denouncing the proposed boycotts to our membership and the media in September 2002 (see www.icsu.org/Media/press/pdf_releases/release1.htm).

Recognizing that the Israeli boycott issue is only one example of the current threats to the universality principle, ICSU is reviewing the principle in the context of the twenty-first century. We are closely monitoring additional calls for boycotts

of Israeli scholars in the French scientific community, as well as security measures that effectively limit the publication of scientific data. We have actively voiced our concern regarding visa regulations that restrict the travel of scientists from specific countries and/or disciplines.

Blakemore *et al.* have highlighted apparent definitional ambiguities in the principle as currently worded, identifying ethical dilemmas (individual and institutional) that warrant consideration. In today’s environment, threats to the principle are as likely to arise from the activities of individuals or small groups as from the policies of a particular government. Clearly, ICSU needs to

develop strategies to address these new circumstances. As we undertake this important task, we welcome input from the broad scientific community. In advance of its meeting, scheduled for March, ICSU’s Standing Committee on Freedom in the Conduct of Science (SCFCS) invites readers to forward additional questions and comments to its chair, Peter Warren, at peter.warren@bee5.demon.co.uk.

Jane Lubchenco (president), Thomas Rosswall (executive director), Peter Warren and all members of the executive board and SCFCS

International Council for Science, 51 boulevard de Montmorency, 75016 Paris, France

Scientific freedom: some face a lonely dilemma

Sir — The Commentary (*Nature* **421**, 314; 2003) by Blakemore *et al.*, on the principle of the universality of science and its obligations on scientists, prompts me to suggest that we should be engaging in a more urgent debate on the role of scientists in legally constituted weapons inspections. The urgency relates to Iraq and the appalling and lonely dilemma that some Iraqi scientists may be facing.

Scientists, like other citizens, have to obey the law of the land, their conscience and the ethical demands of their profession. Many also frame these as religious obligations. Additionally, in the context of disarmament, they have to consider whether they are bound by international laws to which their state either has, or has not, subscribed. They will have loyalties to their country, and may be bound by official-secrets laws. If scientists find any of these demands in conflict, they will encounter naked power and the raw sentiment of fellow-countrymen intolerant of disloyalty.

The approach taken by Blakemore *et al.*, of testing precepts through case histories, is instructive. We could take the case of a Czech defence scientist in 1938. Suppose the scientist knew that the country had not fully dismantled defensive structures as ‘agreed’ in the Munich Treaty. Or what if an individual is engaged in the development of weapons that the country has legally undertaken not to acquire? And if a scientist is ordered to hide from inspectors documents that have international legitimacy, but whose authority runs

counter to national obligations, what action should he or she take?

One option is that of whistle-blowing on the country’s failure to comply with its legal obligations. But should an individual’s opinion on the motives of either regime inform this judgement? If the individual considers the position of both sides to be morally indefensible, would this change the obligation to divulge? Can such a lack of clean hands be allowed to shade one’s willingness to risk liberty and life? And would incompetent concealment be an acceptable moral position or an irresponsible fudge?

Roger Macy

28 Dorking Road, Tunbridge Wells, Kent TN11 2LP, UK

Gene flow might turn wimps into superweeds

Sir — Your News story “Transgenic crop trial’s gene flow turns weeds into wimps” (*Nature* **421**, 462; 2003) highlights the suggestion that gene flow between transgenic crops and potential weeds can act to lessen the latter’s negative effects on important crop plants.

The researchers, Neal Stewart and colleagues, seem to reach this conclusion on the basis of short-term results consisting of a decreased negative effect of initial hybrid weeds on wheat yield when compared to the effects of non-hybrid weeds. Stewart *et al.* apparently attribute this to the reduction of fitness of otherwise well-adapted weeds through the inheritance of genes with high genetic load (that is, deleterious mutations) from the transgenic crop plant.

But the observation that initial hybrids are less aggressive and apparently unable to benefit immediately from such inheritance is not of much importance to agriculturalists. What is important is the evolutionary potential for transgenic crop genes to be shifted and shuffled around in a way that may eventually result in a novel modified gene complex.

Just as Clark Kent was able to change into Superman — as your News story put it — such a novel complex, initially born an inferior weakling, may very well have a chance of becoming a 'superweed'.

Norris Muth

*Department of Ecology and Evolutionary Biology,
University of Tennessee, Knoxville, Tennessee
37996-1610, USA*

Venezuelan government is backing science

Sir — Mendoza *et al.* in Correspondence (*Nature* **421**, 473; 2003) ask for help in the name of the Venezuelan scientific community, in view of the current turmoil in our country. However, as we show here, the real figures tell a different story, pointing to the efforts made by the current government, which has been in power since 1999, to satisfy the demands of the scientific sector.

More than 60% of Venezuela's science budget comes from the government. During the past ten years, total investment has continuously increased, from US\$177 million in 1990 to \$405 million during 2000. In 2001, the state science budget was, astonishingly, increased by more than 50%, reaching \$637 million. As a consequence there have been more grants for established researchers, as well as a significant improvement in economic support for graduate students.

Notably, Venezuela spends more on science and technology activities per researcher than any other country in North or South America apart from the North American Free Trade Agreement (NAFTA) countries: Canada, Mexico and the United States. In 1999, while the NAFTA countries spent \$126,000 per researcher, Venezuela spent \$76,000, compared with an average expenditure of just \$62,000 per researcher in the countries of Latin America and the Caribbean region. In 2000 Venezuela's spending increased to \$86,000 per researcher, while the average in Latin America and the Caribbean went down to \$61,000 (see www.riicyt.edu.ar).

The current government has created a Ministry of Science and Technology and approved the first law promoting science and technology in our country's history.

Regional governments are now required to contribute part of their budgets to science, and a considerable amount of money has been devoted to regional programmes. We believe that the present government has provided benefits to the scientific community, and that its significant efforts for the advancement of science in Venezuela should be recognized.

A considerable number of enthusiastic science students and other young scientists in Venezuela have shown their dedication to science during the current social and political conflict, in particular during December and January, when many of them kept working despite transport difficulties caused by petrol shortages and illegal road blocks.

Despite the efforts made by some political and economic sectors to destabilize the democratically elected government, many members of the scientific community are confident that Venezuelan science will overcome this adverse situation in the near future.

Undoubtedly, though, the current crisis will have a negative effect on this year's science budget (see "Venezuelan science hit by national strikes", www.scidev.net, 7 February 2003).

Juan Luis Cabrera*, **Luis Emilio Guerrero†**, **Arnaldo Donoso***

**Centro de Física, Instituto Venezolano de
Investigaciones Científicas, Apartado 21827,
Caracas 1020-A, Venezuela*

*†Departamento de Física, Universidad
Simón Bolívar, Apartado 89000,
Caracas 1080-A, Venezuela*

*Other signatories to this letter can be found at
<http://caos.fs.usb.ve/carta.html>*

Free access to publicly funded databases is vital

Sir — The International Society for Computational Biology (www.iscb.org) wishes to express regret and concern about the decision by the US Department of Energy (DOE) to shut down the PubScience web site (see *Nature* **411**, 980; 2001 and **418**, 805; 2002). Unfortunately this decision came to our attention too late to comment before the DOE's official deadline passed.

Free access to scientific knowledge and data is essential to scientific progress. Free access to publicly funded databases such as PubScience, PubMed, Medline and GenBank reflects the public's role in funding the science that led to these data, and provides a cost-effective means for disseminating information to the scientific community. It is essential to future progress in scientific research that these public information

resources remain freely accessible.

Philip E. Bourne

*President, International Society for Computational
Biology, San Diego Supercomputer Center,
University of California, San Diego, 9500 Gilman
Drive, La Jolla, California 92093-0505, USA*

GreenSea's interest in fertilizing sea with iron

Sir — Your News feature "The oresmen" (*Nature* **421**, 109–110; 2003) characterizes the objectives of GreenSea Venture as strongly in favour of episodic iron fertilization of selected high-nutrient–low-chlorophyll ocean areas, a non-polluting and inexpensive technique. In fact, we believe only that this may be an efficacious approach to controlling the concentration of carbon dioxide in the atmosphere.

Our activities are in support of 'sound science' and the further development of knowledge applicable to iron fertilization as a carbon-control technology. Only with such knowledge can policy-makers determine whether this technology should be developed, together with other approaches, for efforts at climate control.

M. Lee Rice

*GreenSea Venture, Inc., 20150 Woodtrail Road,
Round Hill, Virginia 20141-1943, USA*

Eastern Europe nurtures talent for the West

Sir — As a scientist from Slovakia, one of the Eastern European countries in the process of joining the European Union, I was pleased to see your Opinion article "Too quiet on the Eastern front" discussing science in Eastern Europe (*Nature* **421**, 459; 2003).

After spending five years doing my PhD in Austria and two more years as a postdoc in England, I was interested in returning to Slovakia to pursue my scientific career there. I was kindly offered a university position with the possibility of starting my own group — but the salary was only SKK100,000–200,000 (\$2,550–\$5,100) a year. Because I have two children I had no choice but to look for other options.

I have realised that, unfortunately, there are no fellowships that would allow me to go back to Slovakia. The situation is similar in all Eastern European countries: universities nurture talents for the Western countries.

Juraj Gregan

*Department of Zoology, University of Oxford,
South Parks Road, Oxford OX1 3PS, UK*

Biodefence on the research agenda

The world needs new and creative ways to counter bioterrorism.

Anthony S. Fauci

The vulnerability of the United States and the rest of the world to attacks of bioterrorism remains starkly exposed by the as-yet unsolved anthrax attacks¹ immediately following the horror of the 11 September 2001 terrorist assault on the World Trade Center. We have invested many resources to control or eradicate pathogenic microbes that, as these events show, cannot now be ignored. It is the responsibility of the biomedical community to undertake research on defence against biological agents with energy, creativity and commitment. The world expects nothing less from us.

The US government is investing an unprecedented amount of money — \$5.9 billion planned for fiscal year 2003 — to counter the threat of bioterrorism. Of that sum, the National Institutes of Health (NIH), the lead government agency in biomedical research, will receive nearly \$1.75 billion, almost eight times the fiscal year 2002 budget for biodefence research, and the largest single increase in resources for any initiative in the history of the NIH. With this largesse come enormous responsibilities.

Preventative steps

The overriding objective of the NIH is to support biodefence research in order to provide people with countermeasures in the form of diagnostics, therapies and vaccines. We at the NIH enthusiastically accept this important new responsibility² and have devised a strategic plan and research agenda through the National Institute of Allergy and Infectious Diseases (NIAID), with major input from the scientific community who will ultimately implement this agenda.

In discussions with the Department of Health and Human Services, the Office of Homeland Security and the US Congress while devising this programme, we underscored the premise that our efforts to defend against bioterrorist attacks would be anchored in the traditional processes of basic biomedical research. We were also acutely aware of the need to rapidly translate basic research results into definable, quantifiable endpoints such as diagnostics, therapeutics and vaccines. A major goal of the NIH has always been the translation of basic research findings into practical interventions, but until now, the path to product development has not been central to our research strategy. Current global threats have compelled us to somewhat modify the way in which we do business.

First, we have increased our commitment to translational research and product develop-

These are lofty goals that may take many years to accomplish.

ment, based on the strongest possible foundation of fundamental knowledge about pathogens and mechanisms of microbial pathogenesis, as well as host responses. We are committed to strengthening basic research endeavours, but it is no longer adequate to pursue an avenue of research, to learn something interesting and potentially important, and then to move on, leaving the translation of that knowledge predominantly in the hands of others. The NIH did not suddenly decide to research potential agents of bioterrorism or appreciate the need to develop practical countermeasures on 11 September 2001, or on the occasion of the anthrax attacks. The NIAID funded research on smallpox, anthrax, Ebola virus, botulinum toxin and other potential bioterror pathogens well before these events. An immediate practical spin-off of this research was the demonstration of the potency of decades-old stores of smallpox vaccine, which when diluted five-fold still generate the skin lesion 'take' regarded as a hallmark of protection³. Such findings provide a readily available countermeasure against a potential bioterror threat at the same time as development of second-generation products proceeds.

Second, we are considering the design and development of biodefence products in new and creative ways. While continuing to develop therapeutics and vaccines targeting specific pathogens, we will devise new strategies for producing broader-spectrum therapeutics and vaccines. The goal of developing 'universal' antibiotics, antivirals and antitoxins, effective against all or most classes of biological pathogens, is not unattainable. Likewise, we will develop safe and effective vaccines against ranges of microbial agents. We will pursue innovative approaches for modulating innate immunity to induce and enhance protection against many biological pathogens, as well as simple and rapid molecularly based diagnostics to detect, characterize and quantify infectious threats. These are lofty goals that may take many years to accomplish — but we must aspire to them.

Third, we must enormously strengthen our interactions with the private sector, including biotechnology companies and large pharmaceutical corporations. Many biodefence-related products that we are

pursuing do not provide sufficient incentives for industry — the potential profit margin for companies is tenuous, and there is no guarantee that products would be used. Therefore, we will seek non-traditional collaborations with industry, for example guaranteeing that products will be purchased if companies sign up (even if the fate of those products is a stockpile), so that we can quickly make available effective vaccines and treatments against agents such as anthrax, botulinum toxin, Ebola and plague.

This concept was mentioned by President Bush in his announcement of the Bioshield project in his State of the Union address on 28 January (see web links).

New strategies

We need a new paradigm to meet the requirements of this sobering international situation within an uncharacteristically brief timeframe for biomedical research. In pursuing basic research, we must never lose sight of the goal of the development of safe and effective countermeasures to protect the public against the threat of bioterrorism. The heightened and urgent need for increased collaborations among the academic sector, government and private industry will surely provide benefits far beyond protection from deliberate acts of bioterrorism. After all, the general philosophy and strategy of bioterrorism defence is essentially the same as that for defence against naturally emerging and re-emerging infectious diseases threatening global public health. The carefully planned use of the new resources will unquestionably have enormous benefits in our struggle against these natural pathogens, as well as other threats to public health that far transcend the spectre of bioterrorism. ■

Anthony S. Fauci is director of the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA.

1. Jernigan D. B. et al. *Emerg. Infect. Dis.* **8**, 1019–1028 (2002).

2. Lane, H. C., La Montagne, J. & Fauci, A. S. *Nature Med.* **7**, 1271–1273 (2001).

3. Frey, S. E. et al. *New Engl. J. Med.* **346**, 1265–1274 (2002).

NIAID strategy on biodefence

► www.niaid.nih.gov/biodefense/research/strat_plan.htm

Status of anthrax investigation

► www.cnn.com/2003/HEALTH/01/30/anthrax.lessons

Bioshield initiative

► www.whitehouse.gov/news/releases/2003/01/20030128-19.html

► www.washingtonpost.com/wp-dyn/articles/A63430-2003Jan29.html

Forever young

Progress in improving health in old age could let people live longer.

Ageless Quest: One Scientist's Search for Genes That Prolong Youth

by Lenny Guarente

Cold Spring Harbor Laboratory Press: 2003.

154 pp. £14, \$19.95

Linda Partridge

Scientists in their later years sometimes enter a philosophypause, characterized by armchair speculations and the writing of autobiographies. *Ageless Quest*, although having this appearance, is an interesting exception, because it is written by a scientist at the peak of his powers. Part autobiography, part social commentary, part science, the book does an excellent job of explaining and extolling the use of model organisms for research into ageing.

The common features of ageing are a decline in fecundity and increased vulnerability to death during adulthood. Many different kinds of damage and pathology accumulate with age, the precise spectrum varying between individuals. The complexity and variability of the ageing process have often led it to be regarded not as one process, but as many. For these reasons, ageing has seemed intractable, both to experimental analysis and to medical amelioration. Although many researchers studying ageing would not sign up to an agenda that aimed to increase lifespan per se, they would agree that their aim is to improve health in old age. The message of *Ageless Quest* is that the prospects for improving health in the elderly are far greater than has been supposed, and that these improvements are, in turn, likely to extend lifespan.

The recent history of scientific thinking about ageing is mirrored in the choices that Lenny Guarente has made in his own research career, a story that he tells with candour and insight. Initially he played safe, focusing his research on the regulation of gene expression, to win tenure at the Massachusetts Institute of Technology, a process akin to dodging a carefully aimed bullet. After tenure and security, he cast around for a riskier, less mature research area, and considered working both on AIDS (too crowded) and the human brain (too difficult). Several straws in the wind suggested that ageing could be amenable to the standard methods of genetic analysis. Slow-ageing, mutant strains of yeast and the nematode worm *Caenorhabditis elegans* had been described. And caloric restriction, in which food intake is reduced to about half of normal levels, had long been known to slow down ageing in rodents and other



Mirror, mirror on the wall: many of us dream of staying young and beautiful forever.

organisms, a phenomenon that is presumably mediated by altered gene expression.

So graduate students Brian Kennedy and Nick Austriaco were given a year to establish the yeast ageing system in Guarente's lab. (An engaging feature of the book is the prominence given to the role of successive graduate students in shaping the science, along with Guarente's thumbnail sketches of their characters.) Kennedy and Austriaco found that different yeast strains aged at characteristically different rates. The team isolated four long-lived mutant yeast strains, one of which was also unable to mate, which allowed the mutation to be pinpointed to a gene whose protein product was already known to be involved in the regulation of gene expression.

The quest to understand how this mutation slows ageing required tenacity, technical ingenuity and considerable thought. It wasn't the mutated gene itself that turned out to be critical, but a partner gene called *SIR2*. The original mutation had resulted in over- rather than under-activity of *SIR2*. The *SIR2* gene therefore acts to promote the survival of the yeast.

SIR2 turned out to have two telling characteristics. First, a homologue in *C. elegans*

also increased lifespan, suggesting that the effect of *SIR2* on ageing is conserved over huge evolutionary distances. Guarente has had what he terms "consistent mentoring" from successive departmental chairmen, and this latest discovery was greeted with: "Just what we need, a long-lived worm." Second, *SIR2* encodes an enzyme (NAD-dependent histone deacetylase) that is involved in chromosomal compaction; when a chromosomal region is condensed, its genes become inaccessible and are turned off. As the enzyme requires NAD, high levels of which may indicate that cellular nutrient supplies are low, it would probably be most active when energy is in short supply. In accordance with this idea, the lifespan of yeast that overexpress *SIR2* is not extended by reduced glucose availability, as it is in wild-type yeast.

Writing about caloric restriction in the late 1980s, David Harrison and Robin Holliday pointed out that mammals and other animals have evolved methods of surviving hard times, particularly food shortage. Starved organisms can suppress processes that lead to reproduction, and enhance processes that protect against environmental stress. These changes may in turn slow down ageing, increasing the likelihood of

successful reproduction when the food supply is restored.

Guarente makes a good case that *SIR2* epitomizes one kind of gene action of the sort that Harrison and Holliday envisaged. It is responsive to cellular nutrient status, and reacts to high NAD levels (and hence food shortage) by shutting down gene activity. Precisely which genes are turned off may differ between organisms. Genes in the insulin/IGF-like pathway, of which *SIR2* may be part, are also nutrient-sensitive regulators of gene expression, and affect the rate of ageing in worm, fruitfly and mouse.

The indications so far from model organisms are that interventions that slow down ageing can also improve age-specific health. Caloric restriction, for example, slows down the accumulation of multiple forms of age-related damage and pathology in rodents.

Guarente is an optimist. He thinks that drugs targeted at products of genes such as *SIR2* will become available in the next 10 or 20 years to improve age-specific health and slow down ageing. Although voluntary caloric restriction is not a feasible approach to improving human health, it may be possible to target the pathway that mediates the response. Indeed, Guarente devotes much of his time to his biotech company Elixir, and has much of interest to say about the synergies and conflicts of interest between academia and business. *Ageless Quest* conveys some quite difficult ideas and complex experimental results with a clarity and freshness that deserve to make it widely read. ■

Linda Partridge is in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

Bonding the two cultures

The Art of Chemistry: Myths, Medicines and Materials

by Arthur Greenberg

Wiley: 2003. 357 pp. £41.95, \$59.95, E66.70

John Emsley

The Art of Chemistry is a companion to Arthur Greenberg's earlier book, *A Chemical History Tour* (Wiley, 2000), and again he offers us a selection from his unique collection of historical material. If you enjoyed the earlier book, which I did, then you will enjoy this as well. It should appeal to chemists, historians and artists, particularly those who are fascinated by alchemy and how this base science was transmuted into the gold of chemistry.

The 71 essays are lavishly illustrated with page after page of reproductions that capture the chemical mood of the time, with Greenberg providing the necessary commentary. A few chapters merely whet the appetite

Base instincts

Alchemy conjures up images of mysterious men in dark laboratories, battling with forces they don't understand in a vain effort to turn base metal into gold. But on what are these images based? *Transmutations: Alchemy in Art* by Lawrence M. Principe and Lloyd DeWitt (Chemical Heritage Foundation, \$25) brings together alchemical paintings from the Eddleman and Fisher collections at the Chemical Heritage Foundation, based in Philadelphia, including *The Search for the Alchemical Formula* by Charles Meer Webb, shown here. This short book provides an introduction to the dark dawn of modern chemistry.



without satisfying it. The one entitled 'Secrets of a Lady Alchemist' concerns Marie Meurdrac, who was the first woman to write a book on chemistry, which was published in 1656, but the coverage is all too brief. So too is 'A Promising President', which gave me a glimpse of unexpected depths to former US president Herbert Hoover. He did a scholarly translation of a 1556 book by Georgius Agricola, *De re metallica* (*Of Things Metallic*), from Latin into English.

While those essays left me hungry for more, they are the exception — most of them provide a meaty meal. I particularly enjoyed the one on William Prout's hypothesis of 1815, in which he theorized that all elements are derived from hydrogen. This theory was seen as demonstrably wrong in the nineteenth century: if hydrogen's atomic weight is 1, how can chlorine's be 35.5? Nevertheless, his hypothesis explained why most atomic weights are integral numbers, and it was indeed prophetic. A century later, chemists learned that an element is identified by the number of protons in its nucleus and, as hydrogen's nucleus is a single proton, it can be seen as the element of the elements.

In a book that is largely pictures, one might expect the accompanying text to be merely an add-on, but *The Art of Chemistry* is well written and peppered with Greenberg's witty comments and cynical asides, providing light relief along the way, although some are more toe-curling than amusing.

Sometimes his remarks require specialist knowledge, such as an understanding of baseball, because he likes to make comparisons between the achievements of great chemists and the giants of his favourite sport.

The more of the book I read, the more I kept on wanting to read, and indeed for me the more interesting essays come in the second half, culminating with the whimsically titled 'Section VIII (Some Fun)' with its articles on occult chemistry and cigarette cards of famous chemists. If your interest is alchemy then the first dozen or so essays in 'Section I (Spiritual and Mythological Roots)' and 'Section II (Still, Cupels, and Weapons)' will be the main attraction.

In the 1950s, C. P. Snow wrote of 'two cultures' and the lack of communication between them, and it really did seem that the worlds of art and science were like oil and water: impossible to mix. They represent different, though equally valid, ways of looking at the world. But that should not prevent artists from producing works with scientific themes, or scientists from appreciating the work of artists. The two worlds have often drawn strength from each other and *The Art of Chemistry* proves that it can be done. I hope that it will encourage those who are currently struggling to meld the two. ■

John Emsley is a science writer in residence at the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK, and is the author of *Nature's Building Blocks*.

Soaking up the limelight

Systema Porifera: A Guide to the Classification of Sponges

edited by John N. A. Hooper &

Rob W. M. van Soest

Kluwer Academic/Plenum: 2002. 1,810 pp.
\$595, £398, €625

Lorraine Berry

Sponges are “among the most successful life forms that have ever existed, with an estimated 15,000 species alive today”, according to the editors of this new work on their classification. The study of sponges has provided information about fundamental biological issues, including the biosynthesis of chemicals, totipotency (the ability of a cell to differentiate into other cell types), the evolution of eukaryotic immunology, and cellular theory. Sponges are also a source of therapeutic drugs, and can serve as biomarkers of pollution and proxies for palaeoclimatic change.

Nomenclature problems have plagued sponge taxonomists for over a century, however. The classification of sponges is complicated by their plasticity of form. They lack recognizable organs, cell types often migrate, and detritus (including skeletal parts called spicules from other sponges) can be incorporated, transforming the sponges beyond recognition. As a result there are many synonyms, which are difficult to reconcile with species descriptions (from different authors at different localities at different times) and with type material that is dispersed among museums around the world.

Systema Porifera attempts to resolve the higher systematics of the phylum Porifera, incorporating spongiomorphs such as the Archaeocyatha, ‘Stromatoporoidea’ and ‘Sphinctozoa’, which previously resided in other phyla. The book represents the work

of 45 researchers, who took 7 years to re-evaluate and define the 2,100 nominal genera. It also provides a review of the taxonomic literature, creating an invaluable database of sponge biodiversity and a platform for the future development of sponge systematics.

Interest in sponges has been growing since the 1960s, when the discovery of bioactive metabolites (used to synthesize compounds with antibiotic and other biomedical properties) caught the attention of drug-development agencies. Since then, the number of recognized species of sponge has doubled, and today they are studied by several hundred researchers in various institutes around the world. This increase has led to an explosion in the discovery and documentation of species, which until now was dispersed in journals and monographs. *Systema Porifera* provides an important service by drawing information from these disparate sources together and updating it.

The editors are well qualified to do this because of their work on the rich sponge faunas of the Indo-Malay Archipelago. John Hooper has also revised the systematics of Australasian sponges, and Rob van Soest has concentrated on sponges from the north-eastern Atlantic and the Caribbean. Together they have done an excellent job, making *Systema Porifera* interesting and accessible to a wider scientific audience than pure sponge taxonomists.

Descriptions of each class, order and family of sponge are organized under a series of uniform subheadings, making information easy to locate. Bibliographies are comprehensive and include many recent reviews. The authors discuss the history of nomenclature and classification, and recognize the possibility of future taxonomic changes, highlighting the fact that sponge science is still evolving.

Most sections contain an identification key for families and genera — something never previously attempted for Porifera —

although I would recommend, as the editors suggest, referring to the *Thesaurus of Sponge Morphology*, edited by Nicole Boury-Esnault and Klaus Rützler (*Smithsonian Contributions to Zoology* 596, 1–55; 1997), to use them. Some diagrams and electron micrographs in *Systema Porifera* are a bit small, although enlarging them would have further increased the size of this already enormous — but invaluable — book.

Lorraine Berry is in the Department of Invertebrate Zoology, Royal Belgian Institute of Natural Sciences, Rue Vautier, Brussels B-1000, Belgium.

Universal building blocks

The Road to Galaxy Formation

by William C. Keel

Springer: 2002. 230 pp. £99.95, \$109, £60

Carlton Baugh

Galaxies are the building blocks of the Universe. Understanding the origin of these vast agglomerations of stars, gas and dust is one of the main challenges facing cosmologists today. Our ideas about the physical processes responsible for shaping galaxies have matured rapidly over the past decade, in response to the staggering advances in astronomical technology over this period.

The advent of 10-metre ground-based telescopes, the success of satellite-borne instruments such as NASA's Hubble Space Telescope, and the opening up of the electromagnetic spectrum far beyond the optical region have revolutionized our view of the Universe. We can now observe galaxies that are so far away that the time it takes light to travel from them is a significant fraction of the age of the Universe. This means that we see these objects as they appeared when the Universe was just a billion or so years old. The task confronting theoreticians is daunting: to explain the formation and evolution of galaxies over more than ten billion years of cosmic history. It is certainly time for a book devoted to setting out the available clues as to how galaxies are made.

William Keel's book outlines the two competing theories of galaxy formation: the ‘monolithic collapse’ and the hierarchical scenario. The former is a throwback to the 1960s, when it was first suggested that galaxies were born in some violent episode in the early Universe. The discovery of galaxies with seemingly elderly stellar populations at high redshifts (which correspond to a time when the Universe was relatively young) gives credence to the monolithic picture, with the result that this simple model has stubbornly refused to die.

The hierarchical scenario, on the other hand, attacks the problem from the opposite



Hanging on: some 15,000 species of sponge, including *Leucetta chagosensis* shown here, are alive today.

direction. In this model, galaxies are built up by repeated mergers of smaller fragments. In this case, there is no single epoch of galaxy formation, but rather a steady build-up of stars. This model is appealing because it is set in the context of a model for the formation of structure in the dark-matter component of the Universe. Models in which the mass of the Universe is dominated by weakly interacting particles known as cold dark matter, and which also have a significant cosmological constant or dark-energy component, are receiving impressive support from measurements of the cosmic microwave background radiation and the Hubble diagram of distant supernovae. However, this evidence is circumstantial: dark matter has not been detected in the laboratory, and there is no convincing theoretical explanation of dark energy.

Keel's own research has covered a wide range of topics, which is reflected in the richness and variety of subjects covered in this book. It is refreshing, in a market dominated by theorists, to come across a book on galaxy formation written from an observational perspective. After all, any model of galaxy formation, no matter how appealing from a theoretical point of view, is eventually judged by how well it describes what is actually out there.

In a book of this length it is impossible to cover in detail all of the most exciting advances of the past few years. Consequently, the discussion of Lyman break galaxies — the first significant population of galaxies to be identified at high redshift — is rather brief in view of the tremendous impact that these objects have had upon the subject. Also, galaxies that are detected by their emission at submillimetre wavelengths (which occurs when the dust that they contain is heated by starlight or by material falling into a central black hole) receive little attention until the final chapter. The book provides a flavour of the physics of galaxy formation, rather than a rigorous review of the theory. Readers who require a comprehensive theoretical briefing could turn instead to *Cosmological Physics* by John Peacock (Cambridge University Press, 1999) or *Cosmology* (2nd edn by Peter Coles & Francesco Lucchin (Wiley, 2002).

Nevertheless, *The Road to Galaxy Formation* should prove to be a handy primer on observations of galaxies for graduate students, advanced undergraduates and theoreticians who feel too shy to visit a telescope. A particularly useful feature is the bibliography at the end of each chapter, which contains a brief résumé of selected research papers and will no doubt be invaluable to newcomers to the field who need guidance in selecting further reading from a burgeoning literature. ■

Carlton Baugh is in the Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK.

Science in culture

Leonardo lifts off

A wing designed by Leonardo da Vinci proves to be aerodynamic.

Martin Kemp

Leonardo da Vinci is famous, among other things, for inventions that did not work, or rather that could not have worked in his own era. He was, as the cliché has it, a man ahead of his time. But like many clichés, this obscures rather than clarifies. Although he was undoubtedly productive as a 'jobbing' engineer, his grand, visionary projects, such as the flying machine and the tank, blended genuine aspirations with bouts of visual boasting directed at prospective patrons.

The brilliance and practical limitations of Leonardo's more extravagant visions have recently been vividly demonstrated in the project to build a flying machine undertaken by ITN Factual for Channel 4 Television in Britain. I was called in as their consultant, to work with Skysport Engineering of Bedfordshire in England, which specializes in the resurrection of early aircraft.

The brief was to construct something that would actually fly, but with the minimum of deviation from Leonardo's own designs, in detail and in spirit. We were determined to use nothing that he did not or could not have envisaged. The great *uccello* (bird) that he envisaged in the 1490s, with its flapping wings, had no chance of flying, so we favoured the gliding mechanisms to which he later resorted.

Our construction was based on a series of Leonardo's bat-like designs, which combined large spans with skeletal lightness (at least in terms of the materials available to him). Leonardo's concept was based on his conviction that nature's own inventions operated with no insufficiency and no redundancy in the context of natural law. Deciding to remake nature on its own terms, he worked on the principle that the aspiring aviator should not literally imitate a bird but should create a mechanical body that works analogously, in conformity with the causes behind natural effects. It is this approach that places the flying machine in much the same category as the *Mona Lisa*, that most synthetically contrived of portraits (*Nature* **389**, 799; 1997).

The problem that we faced, like everyone else who has attempted to fly Leonardo's machines, was that of achieving sufficient lift to raise into the air

the relatively heavy materials available to him. Our breakthrough came with the observation of a detail on a sheet in the Biblioteca Ambrosiana in Milan that had appeared insignificant to my eyes. The Skysport engineers saw it differently. Near the lateral rib adjacent to the pole at the inner margin of the wing, they noticed a looped line that passes over the front edge (circled in red, below). The rear of the loop is labelled "*panno*" (cloth). What this detail said to the engineers was 'leading edge' and 'aerofoil' — the necessary features for lift-off.

However, Leonardo had no sense of the dynamic laws that underlie modern aerofoil design, which involve compression beneath the wing and rarefaction above. During his many hours watching birds wheeling on currents of air, fish swimming in streams, and boats tacking against the wind, he never doubted that water and air (both of which are fluids) behave in the same way. He did not pay attention to the fact that air, unlike water, is compressible, and had not considered such a possibility. But he had observed the robustness of the leading edge of a bird's wing and the relatively broad and blunt heads of some fish, which told him that air or water needed to be pushed strongly out of the way.

The Milanese wing design achieved triumphant lift-off high on the Sussex Downs on a breezy autumn morning about 500 years after it was drawn. In the version that flew, Leonardo's wing design was doubled up, and we omitted the windlass, which was intended to adjust the angle of the wings but would have served no useful purpose. Also added was a leonardesque tail for stability and a somewhat less leonardesque A-frame for the intrepid pilot.

A new formula is needed to replace the cliché. Leonardo, with extraordinary visual inventiveness, envisaged more of the potentiality of his period's science and technology than any of his contemporaries. And, like many highly original inventors, he needed his slice of luck to come up with the right answer for the wrong reasons.

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Winging it: a flying machine based on plans by Leonardo da Vinci has been flown. The red circle shows the loop that indicates the leading edge.



Country life

Summarize yourself in the form of a title of a paper in Nature.

Middle-aged housewife from small-town USA turns to dinosaurs and generates controversy by attempting the impossible.

What was your first experiment as a child?

Burning ants and grasshoppers with a magnifying glass. I was also the subject in a study to discover why people don't rely on a diet of dirt and grasshoppers. The investigator was my brother who fed me strange things that I ate without question because he was my hero.

What makes a good scientific mentor?
Patience!

Whose graduate student would you most like to have been?

Albert Schweitzer, a man who combined medical science with a deep respect for humanity.

Are you by any chance related?

He was a cousin of my ex-husband's father.

What single scientific paper or talk changed your career path?

Peggy Ostrom's presentation on stable isotopes and implications for food webs at the first meeting of the Society of Vertebrate Paleontology I ever attended, in Toronto in 1992.

What literary character would you employ as a postdoc?

Indiana Jones.

What was the worst/most memorable comment you ever received from a referee?

Something along the lines of: "I don't care what the data say, I do not believe molecular preservation is possible in specimens this old."

What book is currently on your bedside table?

Several, as I can't seem to focus on one at a time. The one that comes to mind is about the Civil War — and it's in my car, not by my bed, so I can read it when stuck waiting somewhere.

What music heads the playlist in your car or lab?
Country music, all the way. Is there any other kind?

Where and when would you most like to have lived or worked?

I would love to have joined the battle for Scottish independence, alongside William Wallace or Robert the Bruce, which would mean fighting against my own ancestors, the MacDougalls. From a scientific standpoint, I would love to have been field assistant for Roy Chapman Andrews who led the Trans-Asiatic Expeditions to the Gobi Desert in the 1920s.

The job of captain on the Enterprise in Star Trek has become vacant. Nominate any real person, living or dead, for the post.

Phil Currie of the Royal Tyrrell Museum of Paleontology in Drumheller, Alberta, for his adventurous spirit, curiosity and willingness to take risks. And besides, he is a sci-fi fan.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Jesus Christ — my hero and role model.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Shamelessly eavesdrop. If they like what I do, I will feel honoured; if they criticize it, I might think of questions I had not considered.

What one thing would you rescue from your burning laboratory?

My technician.

What's the best piece of advice you've ever received?

Ignore the critics and do what you know is right. Following this advice is not as easy as it sounds.

What do you do to relax?

In the long Montana winters, I do a lot of reading with my cat on my lap by the fire. In the summers, the more I can get outdoors the more relaxed I am, so I ride horses, hike and bike. And, year round, I try to get in my four-mile daily run.

What would you have become, if not a scientist?

A better mum, perhaps. Professionally, a medical doctor.

What's your motto?

Carpe diem: seize the day — and choke the living daylight out of it.

What previously under-recognized sport should be included in the Olympic Games?

Water ballet.

What's the one thing about science that you wish the public understood better?

I wish people didn't see science as irrelevant or frightening. It is in their best interest to be scientifically literate, so they can make rational decisions rather than risk being swayed by arguments that are better presented emotionally, or better funded, or supported by famous people.

Harry Potter or Lord of the Rings?

Lord of the Rings, by far.



Mary Schweitzer

Mary Schweitzer is a pioneer in the emerging field of studies on biomolecules retrieved from the fossil record. She is in the Department of Microbiology and the Museum of the Rockies at Montana State University, Bozeman, Montana.

You've just been told (in confidence) that the world will end tomorrow. What do you do next?
Go to be near my children ... and pray a lot.

What's the most interesting thing in your fridge?
My room-mate left a watermelon in the refrigerator last summer. I just found it. Diagenesis in action.

What one thing would you change about Nature?
The magazine or the concept?

Touché. The magazine, of course.
I would ask for a more liberal acceptance policy! As far as the concept goes, well, it seems to be working just fine, mostly...

Which actor would best portray you in a film of your life story?
I could hope for Sandra Bullock or Julia Roberts...

Name one extravagance you can now get away with because of your eminence.
Nothing. I got away with more before I ever entered science.

What music would you have played at your funeral?
Country and country gospel music. Songs like *Prop Me Up Beside the Jukebox (If I Die)* come to mind.

What would be the title of your autobiography?
'How on Earth did I end up here?' ■

A fix for RNA

Thomas J. Begley and Leona D. Samson

It has long been known that cells repair chemically or physically damaged DNA. But the discovery that damaged RNA can also be repaired may come as a surprise. What's more, some of the same enzymes are involved.

The 'central dogma' of biology states that DNA makes RNA makes protein, and that this relationship governs the flow of genetic information in almost all living organisms. Damage to any one of these molecules can subvert the normal information flow. Because DNA provides the ultimate repository for genetic information, however, and because DNA damage can lead to permanent sequence changes (mutations), the study of how cells repair damaged DNA has prevailed, to the extent that the repair of damaged RNA (and proteins) has been considered unlikely to even exist. This is in part because, unlike the genome, of which there are just two to four copies per cell, RNAs and proteins are present at tens to thousands of copies and are readily replaced. Damaged RNAs and proteins have thus been considered dispensable.

That view must now be revised, as we learn — from Aas *et al.*¹ on page 859 of this issue — of the repair of damaged RNA *in vivo*. This exciting finding comes hard on the heels of the discovery of a brand new DNA-repair mechanism^{2–4}, catalysed in the bacterium *Escherichia coli* by the enzyme AlkB, and in humans by related proteins. The RNA repair proceeds by the same mechanism.

DNAs, RNAs and proteins have all been reported to be chemically damaged by aberrant methylation — the addition of CH₃ groups. These methyl groups might come from cellular methyl donors such as S-adenosylmethionine, or from environmental pollutants such as tobacco smoke. The consequences of RNA and protein methylation have not been well studied, but aberrant DNA methylation can have profound effects, ranging from the induction of mutations to programmed cell death and concomitant genome degradation (Fig. 1). Fortunately, cells are equipped with several means of repairing DNA that contains methylation damage⁵.

Recently, researchers discovered a new mechanism by which *E. coli* repairs its methylated DNA, namely oxidative demethylation of damaged cytosine (3-methylcytosine) and adenine (1-methyladenine) bases in DNA^{2,3} by the AlkB protein (Fig. 2, overleaf). In working out how AlkB functions, the researchers were helped by a prior computational analysis by Aravind and Koonin⁶ that predicted how AlkB might fold in its active conformation. This analysis identified AlkB

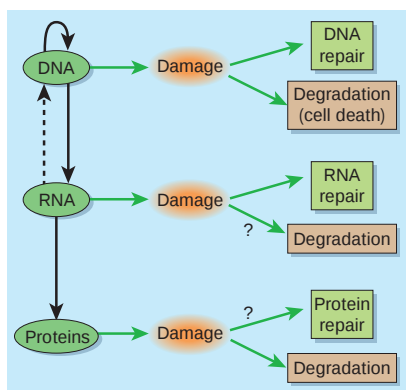


Figure 1 Major biological molecules and the consequences of damage to them. DNA is needed to make RNA, which in turn is needed to make protein, and this relationship governs the flow of genetic information in almost all living organisms. Information can also flow back from RNA to DNA in the case of RNA viruses; and DNA molecules can be replicated in order to propagate life. DNA is the ultimate repository of information, so it is vital that it is repaired when damaged. If it is not repaired it may cause mutations, or it may be degraded as part of a programme of cell death. It has been thought that RNAs and proteins are dispensable, in part because new copies can be easily generated. But Aas *et al.*¹ have found that RNA repair can take place, and may be physiologically important.

as a possible member of the iron-dependent dioxygenase family of enzymes, some of which catalyse oxidative demethylation reactions. Database searches revealed putative AlkB relatives not only in the human genome, but also in the genomes of several RNA viruses, leading Aravind and Koonin to predict that AlkB also acts on methylated RNA. Almost 20 years ago, Karran⁷ reported that methylated RNA bases can act as a substrate for DNA-repair methyltransferases *in vitro*, but the *in vivo* relevance of this was not pursued.

Aas *et al.*¹ now show that Aravind and Koonin were correct in their prediction, and also that RNA repair is indeed relevant *in vivo*. The authors find, remarkably, that bacterial AlkB catalyses oxidative demethylation reactions on chemically methylated RNA as well as DNA. They also provide *in vivo* evidence that AlkB ensures the survival of chemically methylated bacterial viruses that have either RNA-based or DNA-based genomes.

As regards humans, three AlkB relatives have now been identified^{1,4,8} — hABH1, hABH2 and hABH3. Aas *et al.* also show that hABH2 and hABH3 have AlkB-like activity both *in vivo* and *in vitro* (Fig. 2). Biochemically, hABH3 works primarily on single-stranded RNA (and to a certain extent on single-stranded DNA), and hABH2 on single- and double-stranded DNA. The fact that these two enzymes have different subcellular localizations in human cells supports the idea that they have distinct cellular roles. hABH2 is located exclusively in the nucleus; in non-dividing cells it concentrates in nuclear regions called nucleoli, but during the DNA-replication phase of the cell-division cycle it redistributes to sites of DNA synthesis, where DNA transiently becomes single-stranded. hABH3 is located mainly in the nucleus, but also outside it. Its localization in the nucleus is diffuse, and Aas *et al.* presume that it is present at sites at which genes are active, and are being copied into messenger RNAs, in case these mRNAs become damaged.

In essence, then, the activities of these two human proteins add up to that of one bacterial protein, as *E. coli* AlkB repairs both DNA and RNA. It seems that, for each human protein, parameters have evolved to distinguish between RNA and DNA; structural studies will surely shed light on these parameters. Another area for future work will be to generate cells with defects in hABH2-mediated DNA repair or hABH3-mediated RNA repair, in an attempt to identify the relative contributions of each activity in protecting against methylation-induced toxicity.

Why, though, should it be necessary to repair damaged RNA? The answer could be that although DNA is the final arbiter of genetic information, RNA is essential for the most basic biological processes. RNA-based primer sequences are required for DNA replication; and mRNAs, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) are all needed during the elaborate process of protein synthesis. Even the formation of peptide bonds by ribosomes (the cell's protein-making machines) turns out to require catalysis mediated by rRNAs⁹. Moreover, a battery of small, non-protein-coding RNAs regulates a variety of other cellular processes¹⁰.

So maintaining RNA integrity is important for proper cellular function. And repairing damaged RNA may be more efficient

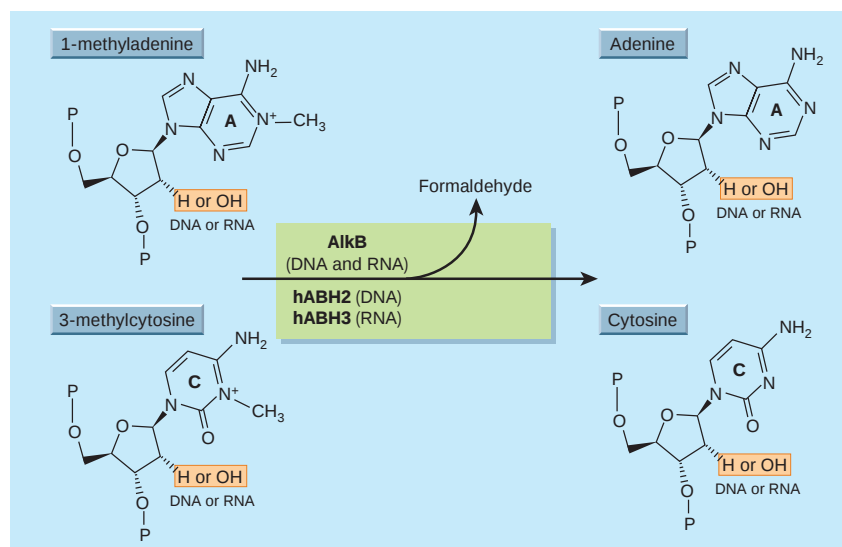


Figure 2 Mechanism for repairing damaged RNA and DNA. Recently^{2,3}, a new mechanism was discovered by which *Escherichia coli* repairs aberrantly methylated adenine and cytosine bases in its DNA. This process, shown here, involves the AlkB protein, which removes the methyl (CH₃) groups. Aas *et al.*¹ have now shown that the same protein can also fix similar errors in RNA. AlkB has three human relatives, one of which (hABH2) can repair DNA, and another (hABH3) RNA. Of the highlighted atoms, H is found in DNA and OH in RNA. Oxygen, iron and 2-oxoglutarate are also required for these reactions.

than destroying it and starting again. Ribosome assembly is a complex, energy-intensive process, and it is not hard to imagine that the thrifty repair of damaged rRNA would be preferable to disassembling or discarding an entire ribosomal particle. Likewise, it takes several hours to produce full-length mRNA copies of large genes, so the repair of damaged copies might make energetic sense. The repair of truncated tRNAs has been documented¹¹, so even the repair of single chemically damaged bases in tRNAs does not seem unreasonable. Most of these RNAs exist as partial duplexes; it will be interesting to see whether AlkB and its relatives repair double-stranded RNA as efficiently as they do the single-stranded variety.

One further question is raised by the fact that many tRNAs and rRNAs have both 1-methyladenine and 3-methylcytosine bases as natural, enzyme-mediated modifications¹². Can AlkB and hABH3 distinguish between these biological modifications and chemical damage? If so, how? And do these enzymes have any role in controlling the normal modification of RNAs? It might be that the RNA-demethylation activity of AlkB-like proteins evolved to regulate biological RNA methylation, and that the repair of aberrant, chemical methylation is fortuitous. Finally, the fact that similar mechanisms are used for both RNA and DNA repair suggests that these nucleic acids might overlap in other ways, with the activation of signal-transduction pathways being a distinct possibility. Damaged DNA is known to signal cell-cycle arrest and programmed cell death. In light of the work by Aas *et al.*¹, we suggest that RNA damage might do the same. ■

Thomas J. Begley and Leona D. Samson are in the Biological Engineering Division, and the Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: lsamson@mit.edu

1. Aas, P. A. *et al.* *Nature* **421**, 859–863 (2003).

2. Falnes, P. O., Johansen, R. F. & Seeberg, E. *Nature* **419**, 178–182 (2002).
3. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. *Nature* **419**, 174–178 (2002).
4. Duncan, T. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 16660–16665 (2002).
5. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (Am. Soc. Microbiol., Washington DC, 1995).
6. Aravind, L. & Koonin, E. V. *Genome Biol.* **2** (3), 1–8 (2001).

7. Karran, P. *Proc. Natl Acad. Sci. USA* **82**, 5285–5289 (1985).
8. Wei, Y. F., Carter, K. C., Wang, R. P. & Shell, B. K. *Nucleic Acids Res.* **24**, 931–937 (1996).
9. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. *Science* **289**, 920–930 (2000).
10. Gottesman, S. *Genes Dev.* **16**, 2829–2842 (2002).
11. Reichert, A. S. & Morl, M. *Nucleic Acids Res.* **28**, 2043–2048 (2000).
12. <http://medlib.med.utah.edu/RNAmods>

Quantum computing

The qubit duet

Gianni Blatter

Small, but consistent, steps are being taken towards the realization of a quantum computer. The demonstration of the coupling of two quantum bits in a solid-state device moves us closer to that goal.

Basic steps towards the creation of a quantum computer have been taken, with the demonstrations of elementary data storage and manipulation using photons and atoms¹ or trapped ions^{2,3} as the quantum bits, or 'qubits'. Solid-state qubits (made, for example, from tiny samples of superconducting material) are an attractive alternative, however, as they could be more easily built into working devices, profiting from the highly developed methods of nanotechnology. So far, solid-state qubits have lagged behind in the race to build the first quantum computer, handicapped by decoherence which implies the breakdown of the quantum information-storing state within the qubit. But new designs have improved the quality of individual qubits, and the next step, to connect qubits together, has now been taken: on page 823 of this issue, Pashkin *et al.*⁴ report the coupling — or rather 'entangling' — of two solid-state qubits.

Quantum computing exploits two resources offered by the laws of quantum mechanics: the principle of superposition

of states and the concept of entanglement. Superposition is a 'one-particle' property; entanglement is a characteristic of two (or more) particles, or, more generally, of two quantum degrees of freedom. Consider a particle with spin — the angular momentum degree of freedom inherent to many elementary particles, such as the electron. With reference to a given axis (say along the *z* direction), the spin of the particle can point in two opposite directions, say 'up' and 'down', and the spin states can be denoted $|\uparrow\rangle$ and $|\downarrow\rangle$. But, by the laws of quantum mechanics, the particle can exist in a superposition of these two states, corresponding to arbitrary rotations away from the axis. The symmetric and antisymmetric states, represented as $[|\uparrow\rangle \pm |\downarrow\rangle]$, are then examples of superpositions in which the spin points along the *x* axis.

Next, consider a two-particle state: there are now four 'basis states', $|\uparrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$, $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$, where the first (second) arrow indicates the state of the first (second) particle. Again, superpositions can be made of these states, including, in particular, the

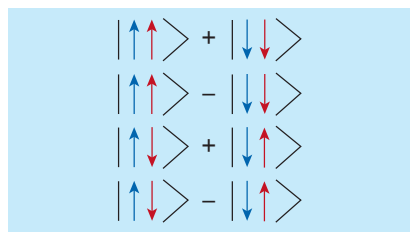


Figure 1 Bell states for a two-particle system. The arrows represent the spin state of each particle (coded red and blue), which may be either 'up' or 'down'.

four 'maximally entangled Bell states', shown in Fig. 1. Such Bell states have the paradoxical property that the particles always 'know' about each other, even if they are separated by huge distances⁵; this property is commonly associated with the 'non-locality' of quantum mechanics. Entanglement such as this is a basic ingredient of quantum computation (Fig. 2) and hence is no longer just a toy for laboratory science^{6,7} but also a tool for the technological world. But how can specific entangled states be constructed on demand?

For a deterministic entanglement of two quantum degrees of freedom, the two constituents must interact in a controlled manner. This has already been achieved in atomic optics, by manipulating a pair of trapped beryllium atoms using a sequence of laser pulses⁸. But now the solid-state camp is catching up: Pashkin *et al.*⁴ have brought two superconducting qubits into a controlled state of interaction.

These qubits encode information using charge rather than spin. In a superconductor, charge is carried by pairs of electrons, known as Cooper pairs. In Pashkin and colleagues' experiment, the qubits are micrometre-size specks of superconductor, each connected to a superconducting 'reservoir'. By quantum mechanical tunnelling, a Cooper pair might jump from the reservoir onto a qubit. Then, analogous to the spin 'up' and 'down' states, the qubit can be in a state with no excess Cooper pairs, or one excess Cooper pair — or a superposed state of both. The superconducting energy gap, which is a natural barrier against the dissociation of Cooper pairs, protects these solid-state qubits from decoherence.

Using lithography and evaporation techniques, Pashkin *et al.* fabricated two qubits on an insulating wafer, connected through a capacitor (see Fig. 1 on page 824). Each qubit coupling to the superconducting reservoir (called a Josephson junction) results in a mixing between the charge states of that qubit, and the capacitive connection between the qubits leads to the mixing of two-particle states, and hence entanglement of the qubit pair. Following the principle demonstrated by Nakamura *et al.*⁹ for a single qubit, Pashkin *et al.*⁴ trace the evolution of the mixing of these two-particle states over

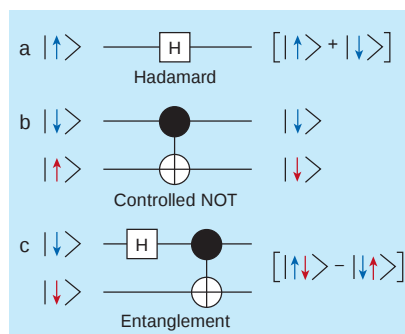


Figure 2 Quantum logic. The quantum state of an individual qubit is described as a complex superposition of two basis states, here denoted by $|\uparrow\rangle$ and $|\downarrow\rangle$. a, Acting on a single qubit, the operation known as 'Hadamard' transforms $|\uparrow\rangle$ or $|\downarrow\rangle$ into the symmetric or antisymmetric superpositions $[|\uparrow\rangle \pm |\downarrow\rangle]$. b, Adding one non-trivial two-qubit operation to the set of single-qubit rotations defines a complete and universal set of gate operations. The operation 'Controlled NOT' flips the target spin (red) if the control spin (blue) points downwards (or else leaves the target unchanged). c, Two-qubit states are complex superpositions of the four basis states $|\uparrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$, $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$. Applying the two operations 'Hadamard' and 'Controlled NOT' to one of the two-qubit basis states produces a maximally entangled Bell state.

time, monitoring the charge oscillations produced with a current probe.

Although they have not yet been able to create and measure a specific entangled state, a numerical analysis shows that the qubit pair does evolve through a maximally entangled state, bringing us closer to the construction of a solid-state quantum logic gate to produce deterministic entanglement. The next step could be to introduce time-controlled switching of the interaction and subsequent readout, measuring the degree of

entanglement through the amplitudes of the final state expressed in the Bell-state basis¹⁰.

Astounding progress has been made over the past few years in the experimental development of quantum computing. In quantum atom optics, entanglement of four particles has been achieved¹¹ and entangled ions are being used to test quantum mechanical relations known as Bell's inequalities¹². Shor's factorization algorithm has been implemented¹³ with a seven-qubit molecule using NMR techniques, successfully identifying the factors of 15 as 3 and 5. From their first appearance a few years ago^{9,14,15}, superconducting qubits have already matured into effective and efficient devices^{16,17}, and now Pashkin *et al.*⁴ have demonstrated the first step towards their deterministic entanglement. These are firm foundations for the development of a solid-state quantum processor, but considerable effort will be needed to direct the complex choreography of a real quantum algorithm. ■

Gianni Blatter is at the Institute of Theoretical Physics, ETH-Hönggerberg, CH-8093 Zürich, Switzerland.

e-mail: blatterj@phys.ethz.ch

1. Raimond, J. M., Brune, M. & Haroche, S. *Rev. Mod. Phys.* **73**, 565–582 (2001).
2. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
3. Gulde, S. *et al. Nature* **421**, 48–50 (2003).
4. Pashkin, Yu. A. *et al. Nature* **421**, 823–826 (2003).
5. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47**, 777–780 (1935).
6. Glauber, J. F. & Shimony, A. *Rep. Prog. Phys.* **41**, 1883–1927 (1978).
7. Aspect, A., Grangier, P. & Roger, G. *Phys. Rev. Lett.* **49**, 91–94 (1982).
8. Turchette, Q. A. *et al. Phys. Rev. Lett.* **81**, 3631–3634 (1998).
9. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. *Nature* **398**, 786–788 (1999).
10. Hill, S. & Wootters, W. K. *Phys. Rev. Lett.* **78**, 5022–5025 (1997).
11. Sackett, C. A. *et al. Nature* **404**, 256–259 (2000).
12. Rowe, M. A. *et al. Nature* **409**, 791–794 (2001).
13. Vandersypen, L. M. K. *et al. Nature* **414**, 883–887 (2001).
14. Friedman, J. R. *et al. Nature* **406**, 43–46 (2000).
15. van der Wal, C. H. *et al. Science* **290**, 773–777 (2000).
16. Vion, D. *et al. Science* **296**, 886–889 (2002).
17. Yu, Y., Han, S., Chu, X., Chu, S.-I. & Wang, Z. *Science* **296**, 889–892 (2002).

Neurobiology

Interneurons take charge

Edvard I. Moser

The brain's hippocampal region contains many classes of interneurons, which, it transpires, show different patterns of activity. They might contribute to memory by shaping the dynamics of neuronal networks.

Learning something new — and recalling what you've learned — involves large numbers of nerve cells, distributed throughout different parts of the brain. Neuronal assemblies in the hippocampus, for instance, are thought to be essential in encoding, consolidating and retrieving memories. These neuronal networks alternate between several functional states, each characterized by a temporally structured

pattern of electrical activity that may facilitate a particular aspect of memory processing. On page 844 of this issue, Klausberger *et al.*¹ present results that bring us closer to understanding the precise but varied ways in which distinct classes of hippocampal neurons impose such network states. Their study focuses on interneurons — small neurons that are involved in the local processing of nerve signals. They find that these

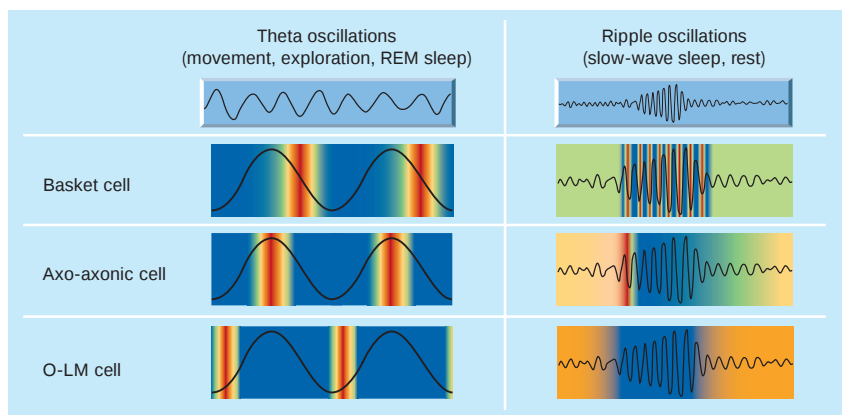


Figure 1 Interneurons and electrical oscillations. The figure shows the activity profiles of three types of hippocampal interneuron during two brain states, based on the findings of Klausberger *et al.*¹. Colours indicate the probability that a given interneuron will fire (maximum red, minimum blue). During theta oscillations, basket cells fire on the descending phase of local theta waves, axo-axonic cells fire just after the peak, and O-LM cells fire at the trough. During ripple oscillations, basket cells discharge one or several phase-locked spikes, axo-axonic cells fire only at the beginning of the ripple sequence, and O-LM cells become silent. The variation within each group is small, suggesting that classes of interneurons exert precise control over distinct aspects of hippocampal network dynamics.

cells, which generally have inhibitory activities, seem to work together in a strictly coordinated manner to control assemblies of excitatory pyramidal nerve cells.

In particular, the new results¹ show that the dynamics of hippocampal networks are related to the diversity in the hippocampal interneuron population. Pyramidal cells are relatively uniform in their structure and behaviour. But interneurons form distinct classes, in terms of their shape, the inputs to which they respond, the other neuron populations they connect to, and the specific parts of neurons with which they make contact^{2–4}. Klausberger *et al.* now show that morphologically distinct classes of hippocampal interneurons also contribute differently to network states.

The authors recorded the electrical impulses (spikes) generated by three types of hippocampal interneurons — basket cells, axo-axonic cells, and oriens–lacunosum–moleculare (O-LM) cells — during two network states in anaesthetized rats. One of these states is characterized by theta oscillations (which have a frequency of 4–10 Hz), the other by short ripple oscillations (120–200 Hz) on a background of more irregular activity. In conscious animals, theta oscillations are associated with movement and exploration, whereas ripples are associated with inactivity and slow-wave sleep⁵.

The three classes of interneurons exhibited distinct, state-dependent patterns of activity (Fig. 1). For instance, during theta oscillations, basket cells fired on the descending phase of each wave, axo-axonic cells just after the peak, and O-LM interneurons at the trough. During ripple oscillations, basket cells discharged one or more phase-locked spikes; axo-axonic

cells fired at the beginning of each ripple sequence; and O-LM cells became totally silent. The implication is that the three classes of interneurons, through their characteristic patterns of activity, make specific contributions to the production of hippocampal network states.

It remains to be seen, however, exactly what these contributions are. For example, what is the function of the sudden drop in activity of O-LM cells during ripple oscillations? As interneurons generally inhibit other nerve cells, this drop in activity means a drop in inhibition. Axons from O-LM cells target the outermost portion of pyramidal-cell dendrites — the area in which pyramidal cells receive excitatory inputs from another brain region, the entorhinal cortex, which mediates highly processed sensory information to the hippocampus. So does the drop in inhibition amplify the cortical input? Does it enable the excitatory synapses (connections) between entorhinal cells and hippocampal cells to become modified, and do such changes occur specifically during ripple oscillations? During ripples, pyramidal cells tend to fire in patterns that are reminiscent of their firing during recent awake behaviour⁶. Such 'reactivation' may contribute to memory consolidation in the hippocampus and the neocortex⁷, where the permanent storage of memories is thought to take place⁷. Are O-LM cells involved in this process?

Moreover, what is the function of the precisely timed firing of basket cells and axo-axonic cells during ripple and theta oscillations? Does it synchronize the output from hippocampal pyramidal cells to the neocortex⁸? And, if so, how do the neocortical target neurons respond to such output?

To determine how the interneurons' precise control of network dynamics contributes to hippocampal functions such as memory processing⁹, the analysis must be extended to behavioural studies in conscious animals. Functional diversity in the hippocampal networks of conscious rats was first studied 30 years ago¹⁰, when nerve impulses recorded in the hippocampus were observed to originate from either 'complex-spike cells' or 'theta cells'. The complex-spike cells discharged at low rates but in bursts; theta cells were more active but their spikes were also more dispersed. These two types of neurons had the properties expected of pyramidal cells and interneurons, respectively. Their identity was eventually verified by staining the recorded neurons in anaesthetized rats¹¹; this confirmed that spike parameters and spike patterns can indeed predict the morphology of hippocampal neurons.

The results of Klausberger *et al.*¹ show that a similar approach can be used to distinguish between classes of interneurons. The spike activity generated by interneurons in a given class varies little, suggesting that state-dependent activity patterns provide valid signatures of basket cells, axo-axonic cells and O-LM cells in anaesthetized rats. Can these differences be extrapolated to awake animals? Similar profiles of activity have been observed during corresponding network states in anatomically unidentified hippocampal interneurons in conscious rats¹². Obviously, there are differences between the awake and anaesthetized conditions¹³, but the new results represent a first step towards a physiological classification scheme that could be used to relate variations in activity profiles to particular interneurons and to specific memory operations¹⁴.

For these results to be applied to behavioural studies, other types of hippocampal interneurons must be analysed in a similar way. The criteria that distinguish between basket cells, axo-axonic cells and O-LM cells in the study by Klausberger *et al.* might be less useful in samples that also contain other interneuron classes. It may be necessary to explore additional parameters, such as waveform shape — which is known to distinguish pyramidal cells from some interneurons^{10,11} — as well as other network states and the cells' responses to specific drugs.

A physiologically based classification system for analysing behaviour represents one of two developments that should advance our understanding of hippocampal interneurons. The other is the genetic manipulation of mice, which may soon allow scientists to examine neuronal networks and memory processes when the function of particular interneurons has been genetically altered. Together, these new tools have the power to uncover the most fundamental

principles of memory formation in the neuronal assemblies of the hippocampus. ■
 Edvard I. Moser is at the Centre for the Biology of Memory, Norwegian University of Science and Technology, MTFs, 7489 Trondheim, Norway.
 e-mail: edvard.moser@cbm.ntnu.no

- Klausberger, T. *et al.* *Nature* **421**, 844–848 (2003).
- Buhl, E. H., Halasy, K. & Somogyi, P. *Nature* **368**, 823–828 (1994).
- Freund, T. F. & Buzsáki, G. *Hippocampus* **6**, 347–470 (1996).
- McBain, C. J. & Fisahn, A. *Nature Rev. Neurosci.* **2**, 11–23 (2001).
- Buzsáki, G. *Neuroscience* **31**, 551–570 (1989).
- Wilson, M. A. & McNaughton, B. L. *Science* **265**, 676–679 (1994).
- Squire, L. R. & Alvarez, P. *Curr. Opin. Neurobiol.* **5**, 169–177 (1995).
- Cobb, S. D., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. *Nature* **378**, 75–78 (1995).
- Paulsen, O. & Moser, E. I. *Trends Neurosci.* **21**, 273–278 (1998).
- Ranck, J. B. Jr *Exp. Neurol.* **41**, 461–531 (1973).
- Henze, D. A. *et al.* *J. Neurophysiol.* **84**, 390–400 (2000).
- Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A. & Buzsáki, G. *J. Neurosci.* **19**, 274–287 (1999).
- Fox, S. E. & Ranck, J. B. Jr *Exp. Brain Res.* **62**, 495–508 (1986).
- Fyhn, M., Molden, S., Hollup, S., Moser, M. B. & Moser, E. I. *Neuron* **35**, 555–566 (2002).

Animal behaviour

How self-organization evolves

P. Kirk Visscher

Self-organized systems can evolve by small parameter shifts that produce large changes in outcome. Concepts from mathematical ecology show how the way swarming bees dance helps to achieve unanimous decisions.

Work published in *Proceedings of the Royal Society*, Mary Myerscough¹ has taken a novel approach to the modelling of group decision-making by honeybee swarms when they are in search of a new home. Bees ‘waggle dance’ to communicate locations of food in foraging, and of potential nest sites when a colony moves during swarming. Myerscough treats the scout bees dancing for alternative sites as populations, and models their growth and extinction with the tools of mathematical ecology. From this approach it is evident how a slight difference in the way the dance-language ‘recruitment’ of other bees is structured in foraging and house-hunting influences the outcome of each process.

The choice of a new home site by a swarm of honeybees is a striking example of group decision-making. When a swarm clusters after leaving its natal colony (Fig. 1), scouts search the countryside for cavities with the appropriate volume and other characteristics². They then return to the swarm, and communicate the distance to and direction of the sites that they have found with waggle dances³, just like those used for communicating locations of food sources in foraging⁴. Usually, the scouts find and report several sites, but in time dances cease for all but one of them, and finally the swarm flies to the selected cavity. Self-organizing processes such as this, in which a complex higher-order pattern (here, the development of a consensus on the best site) arises from relatively simple responses of individuals with no global view of the situation, are receiving increasing attention as biological mechanisms for elaborating complexity⁵.

The population-biology metaphor is appropriate for analysing honeybee dance information. Bees recruited by dances for



Figure 1 Honeybee swarm in search of a new nest site.

a particular site may visit it and in turn dance for new recruits, so dances reproduce. But nest-site scouts may cease dancing before they recruit at least one other dancer: the population of dancers for that site then declines, and may become extinct. Myerscough's approach incorporates key aspects of the dynamics of nest-site recruitment, and can accommodate differences that are specific to the nest site or the individual bee. The populations of dancers have ‘age structure’ in the sense that some dances are a scout's first dance for a nest site, others follow a second trip, and so on. This is similar to population growth with discrete generations, which can be represented in a standard tool of mathematical ecology: a Leslie matrix. The ‘age structure’ patterns also can incorporate an important difference in dance language use between nectar foraging and house-hunting. In foraging, the number of waggle runs that a bee performs when returning with food increases and then levels off with successive dances by that bee (Fig. 2a). In contrast, in house-hunting, the number of waggle runs (which initially depends on the quality of the site) generally declines with each successive dance (Fig. 2b),

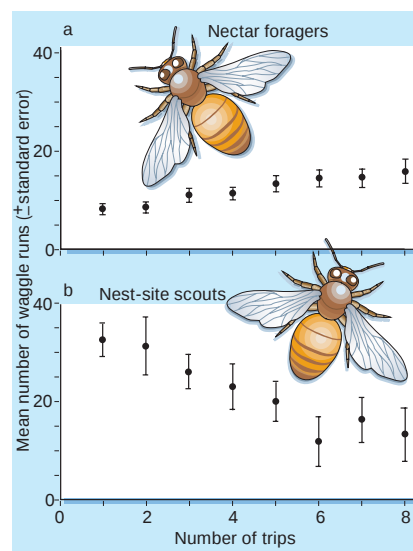


Figure 2 Different patterns of dance-language performance in nectar foragers and nest-site scouts. These graphs plot the number of waggle runs in the recruitment dances performed after each return trip to the colony for successive instances where each individual bee danced¹⁰. a, Nectar foragers continue to dance for many trips. (Here, 93% of 40 foraging bees in 3 colonies danced on more than 8 trips; most danced on more than 50 trips.) b, Nest-site scouts, searching for a new home following swarming, perform dances with more waggle runs at first, but soon cease to dance entirely. (Here, fewer than 5% of 86 bees in 3 swarms performed more than 8 dances.) Myerscough's analysis¹ suggests that this difference in dance performance underlies the difference in outcome: in foraging, it is desirable to recruit new foragers for several sites; in swarming, unanimity for a single site must be reached.

and each scout soon ceases dancing entirely. This gives different patterns of ‘age-specific fecundity’ to the dancing bee populations.

Because the mathematical theory of models of this type is well developed, Myerscough's approach has an analytical payoff. It is straightforward to predict whether a population of dancers for a site will increase or decline. But this is a dynamic process, because only a limited number of scouts can be recruited. As a result, whether dancers for a particular site increase or decrease in number depends both on the quality of the site and on the populations of other dancers. The dancing for a site may increase while competing dances are rare, but then decline in favour of other sites with greater ‘fecundity’ (that is, those that elicit a greater number of waggle runs of dancing per trip by scouts). Such dynamics are typical of swarms^{3,6,7}, with the outcome that the highest-quality site among those discovered is usually selected⁸.

The most striking result of this approach is that it shows how certain special features of the dance in the context of house-hunting

ensure that one, and usually only one, of the populations of nest-site dancer ends up with all available recruits. This finding is of wide interest, because it shows how natural selection can shape a self-organizing process. In both foraging and nest-site scouting, global patterns of allocation of bees among alternative resources arise from interactions of bees responding to their own experience, without a global view of the pattern of allocation or direct knowledge of the characteristics of alternative sources⁹.

However, the contexts of nectar foraging and nest-site decision-making differ in one key respect. In foraging it is usually desirable for the bee colony to use several food sources simultaneously, especially if they are similar in quality; in house-hunting the colony has to settle on just one of multiple sites, even if they differ little in quality. The dance language is used to recruit bees in both settings, but certain aspects of how the dance is performed are different. Myerscough shows it is just these parameters that determine the outcome. Attrition in dances in the Leslie matrix models mathematically ensures that one resource will always dominate in nest-site selection (unless stochastic differences intervene, which may account for the occasional failure of swarms to achieve unanimity). But

in foraging there is no advantage to doing this, and attrition does not occur.

A common misconception about self-organization in biological systems is that it represents an alternative to natural selection⁵. This example illustrates how natural selection presumably evolves such mechanisms: slight modifications of key components shape the parameters of the self-organizing system, and shift the ensuing large-scale patterns to achieve different ends. ■

P. Kirk Visscher is in the Department of Entomology, University of California, Riverside, California 92521, USA.

e-mail: visscher@mail.ucr.edu

1. Myerscough, M. *Proc. R. Soc. Lond. B* published online 3 February 2003 (doi:10.1098/rspb.2002.2293).
2. Seeley, T. D. & Morse, R. A. *Behav. Ecol. Sociobiol.* **49**, 416–427 (1978).
3. Lindauer, M. Z. *Vergl. Physiol.* **37**, 263–324 (1955).
4. von Frisch, K. *The Dance Language and Orientation of Honeybees* (Harvard Univ. Press, 1967).
5. Camazine, S. *et al.* *Self-Organization in Biological Systems* (Princeton Univ. Press, 2001).
6. Seeley, T. D. & Buhrmann, S. C. *Behav. Ecol. Sociobiol.* **45**, 19–31 (1999).
7. Camazine, S., Visscher, P. K., Finley, J. & Vetter, R. S. *Insectes Soc.* **46**, 348–360 (1999).
8. Seeley, T. D. & Buhrman, S. C. *Behav. Ecol. Sociobiol.* **49**, 416–427 (2001).
9. Camazine, S. & Sneyd, J. J. *Theor. Biol.* **149**, 547–571 (1991).
10. Beering, M. A. *A Comparison of the Patterns of Dance Language Behavior in House-hunting and Nectar-foraging Honey Bees* (*Apis mellifera* L.). Thesis, Univ. California, Riverside (2001).

Electronics

Polymers light the way

Andrew Holmes

Using the methods of polymer deposition that are employed in making integrated circuits, light-emitting polymers can be patterned for application in flat-screen, full-colour displays.

Liquid-crystal devices dominate the market for the flat-panel displays used in laptops, personal organizers and mobile telephones. They have their drawbacks, however, and light-emitting polymers are showing great promise as a complementary technology. Processing such polymers to produce a colour, pixelated display is one of the challenges. As they describe on page 829 of this issue, it is a challenge that Müller *et al.*¹ have tackled in a new way.

The disadvantage of liquid-crystal devices is that the light must pass through various colour and polarizing filters before it reaches the eye. So, as everyone knows who has travelled in an aircraft with personal video screens, they can be viewed conveniently only if the screen is at right angles to the viewer. For flat-panel screens, one solution is to use organic fluorescent materials, which are themselves the actual light source and in principle visible from a much wider range of viewing angles. The emissive material can be a thin film of either an organic molecule or a polymer; fluorescence (electrolumin-

escence) is induced by the injection of charge into a film of the emitter sandwiched between oppositely charged electrodes (ideally) powered by a small battery. Good red, green and blue electroluminescent materials are now available, and car radio and multicolour mobile telephone displays using small-molecule 'organic light-emitting diodes' (OLEDs) are on the market². The drawback is that such materials can only be deposited using vacuum (sublimation) deposition techniques, in combination with a shadow mask to control where the molecule is deposited. This presents a problem of scale in large-area displays, although prototype television screens have been fabricated.

By contrast, fluorescent polymer light-emitting diodes (PolyLEDs) can be assembled by deposition from solution. Here the problems are to avoid impurities (in the polymer and the solvent) and not to dissolve away a film during deposition of another layer. One elegant method of delivering a polymer droplet of the right colour to a small dot (pixel) in the display is to use ink-jet

printing³, and rapid progress has been made towards television-size prototype displays using ink-jet printing onto specially prepared wafers of polysilicon (Fig. 1). Simple monochrome PolyLED products are now also on the market, as demonstrated by the display in the electric shaver used by Pierce Brosnan in the latest James Bond movie *Die Another Day*. Müller *et al.*¹ now describe a completely different way of solution-processing coloured displays, one that involves a clever chemical crosslinking method.

Electroluminescent devices operate by forcing positive and negative charges from oppositely charged electrodes into a sandwich device containing a thin film of the fluorescent organic or polymeric material⁴. The charges migrate in opposite directions through the material until they annihilate and cause fluorescence from the excited state. One of the most powerful families of stable light-emitting polymers is the polyfluorenes, which can conveniently be prepared in good yield and high molecular weight by the Suzuki reaction. Generically, this involves carbon-bond formation between aryl halide and boron compounds. In the case of producing polyfluorenes, it is the palladium-mediated polycondensation of a bis-boronate ester with an appropriate dibromo-substituted aromatic compound⁵.

The reaction schemes used by Müller *et al.* are outlined in Fig. 1 of their paper on page 830. They obtained the three primary polymers (red, green and blue) by 'tuning' the Suzuki copolymerization^{6,7} of the bis-boronate monomer with the comonomer containing reactive oxetane end-groups and various dibromo-substituted aromatic comonomers. To form a patterned device, each polymer was crosslinked using the standard photoresist techniques that are employed to make integrated-circuit patterns on silicon chips. Thus, solution deposition of the first polymer onto a transparent electrode (precoated with a conducting polythiophene derivative) in the presence of the photo-acid generator, followed by irradiation of the film through a shadow mask (diffraction grating), released photochemically generated acid in the regions under irradiation.

The acid released in the film caused the strained-ring oxetane end-group to undergo a ring-opening cationic polymerization, leading to crosslinked material. Washing with solvent removed the material that had not become crosslinked, and further gentle baking left the polymer in a well-defined pattern. The two remaining layers of emissive polymers were then deposited in the same way, followed by vacuum deposition of the top electrode, to give a device that showed good resolution and characteristics.

It might have been expected that release of acid and crosslinking would adversely affect the performance of the light-emitting



Figure 1 An ink-jet show. This full-colour, light-emitting polymer display, printed by the ink-jet technique using polyfluorenes, is 2.8 inches in size, diagonally. It has 100 dots per inch, and displays 64 colours plus grey scale. This is the current standard for solution-processing technology. (Picture courtesy of the Seiko-Epson Corp.)

device. Certainly, some of the polymers that were crosslinked showed reduced efficiency compared with those that were not, but others shone more brightly and could be operated at higher 'drive' voltages, apparently as a result of crosslinking. All of the crosslinked devices exhibited lower 'turn-on' voltages, which may have been a sign of the 'protonic doping' that often reduces the barriers to

charge injection at the electrode interface. In general it seems likely that some of the other, more poorly performing features will be improved when devices are optimized.

This work is an exciting demonstration of the possibilities that can emerge from collaboration between polymer chemists and device physicists, chemists and engineers. But it remains to be seen whether devices patterned in this way can match the operating lifetimes of more than 10,000 hours exhibited by their ink-jet-printed counterparts, and whether their operation will be affected by the residual photo-acid residing in the polymer films. ■

Andrew Holmes is in the Melville Laboratory, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.
e-mail: abh1@cam.ac.uk

1. Müller, D. C. *et al.* *Nature* **421**, 829–833 (2003).
2. Shaw, J. M. & Seidler, P. F. *IBM J. Res. Dev.* **45**, 3–9 (2001).
3. Heibner, T. R., Wu, C. C., Marcy, D., Lu, M. H. & Sturm, J. C. *Appl. Phys. Lett.* **72**, 519–521 (1998).
4. Kraft, A., Grimsdale, A. C. & Holmes, A. B. *Angew. Chem. Int. Edn Engl.* **37**, 402–428 (1998).
5. Scherf, U. & List, E. J. W. *Adv. Mater.* **14**, 477–487 (2002).
6. Bernius, M. T., Inbasekaran, M., O'Brian, J. & Wu, W. *Adv. Mater.* **12**, 1737 (2000).
7. Rees, I. D., Robinson, K. L., Holmes, A. B., Towns, C. R. & O'Dell, R. *MRS Bull.* **27**, 451–455 (2002).

Molecular chaperones

Plugging the transport gap

R. John Ellis

Molecular chaperones are generally thought to protect newly synthesized proteins and ensure that they fold into the correct shape. But it seems that two chaperones also help to target certain proteins to mitochondria.

Mitochondria are the major energy-producing organelles of animal and fungal cells, and contain more than 700 different proteins. Fewer than 5% of these proteins are encoded by mitochondrial genes and produced inside mitochondria. Instead, most are encoded by genes in the cell nucleus. They are synthesized as preproteins in the cytoplasm by complex molecular machines called ribosomes, before being transported into developing mitochondria. Writing in *Cell*, Young *et al.*¹ report that the transport of certain preproteins into mitochondria requires the chaperone proteins Hsp70 and Hsp90. The chaperones act by binding preproteins to a specific structure in a receptor protein called Tom70, which lies in the outer mitochondrial membrane (Fig. 1, overleaf). This paper helps to plug the gap left by two reports^{2,3} from the early years of 'chaperonology', which indicated that Hsp70 is involved in transporting proteins

into mitochondria, although the transport mechanism was unknown. It also provides the first evidence that Hsp90 is involved in targeting proteins to a specific organelle.

Mitochondrial preproteins face several problems. After they have been released from ribosomes, they must be targeted to mitochondria but not to other organelles in the cell. They must also adopt conformations that can unfold easily, because they move as extended chains through the two membranes (outer and inner) that surround mitochondria. Such incompletely folded proteins often display hydrophobic regions on their surface, especially if they are membrane proteins, creating the danger of self-aggregation. These problems are combated by molecular chaperones, a large class of proteins that assist the correct folding and assembly of other proteins without remaining bound to them once they are performing their normal functions⁴.

The targeting problem is solved in part by the fact that preproteins contain one of two types of amino-acid sequence that enable them to be recognized by one of two types of receptor in the outer mitochondrial membrane⁵. These are called Tom receptors — short for 'translocase of the outer mitochondrial membrane'. The Tom20 receptor binds preferentially to 'presequences' found at the amino terminus of some preproteins; presequences form positively charged helices that are removed after transport through the mitochondrial membrane, and so do not appear in the mature protein. In contrast, the Tom70 receptor binds to internal sequences found in membrane proteins such as the phosphate and adenine-nucleotide carrier proteins of the inner mitochondrial membrane. Although such sequences are not removed after transport, these carriers are also called preproteins at the stage before transport.

It now seems, however, that this sequence information in preproteins is not the only means of targeting them to mitochondria; their chaperones also have a part to play, at least in the case of membrane preproteins¹. Previous work had shown that a portion of Tom70 located just outside the mitochondrion (that is, in the cell cytosol) recognizes the internal targeting sequences of membrane preproteins⁶. This cytosolic portion also contains several so-called tetratricopeptide repeat (TPR) sequences, which form a characteristic structure within a protein that consists of a helix, then a turn, then a helix. Three of Tom70's TPR sequences are rather similar to sequences in a class of proteins that do not bind preproteins themselves, but instead act as cochaperones by binding to the chaperones Hsp70 or Hsp90. The TPR sequences in the cochaperones form what is called a dicarboxylate clamp domain, which interacts with the aspartate amino acid at the carboxy-terminal end of Hsp70 and Hsp90. Specificity for either Hsp70 or Hsp90 is determined by hydrophobic contacts with neighbouring amino acids in the chaperones.

Young *et al.*¹ now build on these findings. They first suggest that, given the similarity between the TPR motifs of Tom70 and those of the cochaperones, these sequences in Tom70 also form a clamp domain. They further propose that this domain contributes to preprotein targeting by binding the chaperones.

The authors find that the cytosolic segment of human Tom70 binds both Hsp90 and Hsp70 in crude cell extracts, but cannot do so if a key arginine amino acid in Tom70's presumed clamp domain is mutated to alanine. Interestingly, yeast Tom70 binds to Hsp70 only, but this interaction is also abolished by the clamp mutation. This difference between yeast and mammals is due to the receptor, not the chaperones. As an

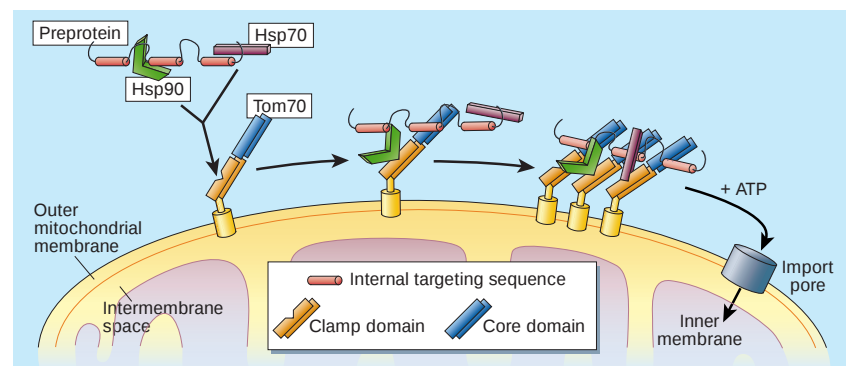


Figure 1 Targeting proteins to mitochondria, as proposed by Young *et al.*¹. Newly synthesized preproteins that are destined to function in mitochondrial membranes are imported via the Tom70 receptor. From left to right: in mammals, the chaperones Hsp90 and Hsp70 bind to the preprotein in the cytosol (in yeast, just Hsp70 is required). The chaperones dock onto Tom70's clamp domain, allowing internal targeting sequences in the preprotein to be recognized by Tom70's core domain. Further Tom70 proteins are then recruited. ATP hydrolysis, probably by the chaperones, is needed to transfer the preproteins from Tom70, through the import pore in the outer mitochondrial membrane, to import machinery in the inner membrane. (Modified from ref. 1.)

indication of the importance of Tom70's clamp domain to mitochondrial protein import, yeast cells die if the normal Tom70 receptor is replaced by one with the clamp mutation. Moreover, import of the phosphate-carrier protein by isolated rat liver mitochondria is inhibited by adding surplus Hsp90 carboxy-terminal domain, which competes with the complete Hsp90 for binding to Tom70's clamp domain. The import of preproteins that use the Tom20 receptor is not affected. Similar experiments with yeast mitochondria indicate that Hsp70–Tom interactions are also essential for preprotein import via Tom70.

Young *et al.* also start to disentangle the respective roles of Hsp90 and Hsp70. Both chaperones have the ability to hydrolyse ATP molecules, and the turnover of ATP causes release of bound preprotein. The ATP-hydrolysing activity of Hsp90, unlike that of Hsp70, is inhibited by the antibiotic geldanamycin, providing a means of testing the specific involvement of Hsp90 in protein transport. Young *et al.* find that geldanamycin inhibits the import of the phosphate-carrier protein both into isolated rat liver mitochondria and into mitochondria of a mammalian cell line. As might be expected, it does not affect the import of the adenine-nucleotide carrier into isolated yeast mitochondria. Moreover, the carboxy-terminal domain of another cochaperone, Bag-1, binds to the ATP-hydrolysing site of Hsp70 but not to that of Hsp90. Surplus Bag-1 carboxy-terminal domain inhibits the import of the two carrier preproteins into mitochondria from either rat liver or yeast by interfering with the turnover of ATP by Hsp70.

These observations suggest that Tom70 is not only a preprotein receptor, but also a cochaperone that aids chaperones in the targeting of carrier preproteins to mitochondria (Fig. 1). This dual function of

Tom70 can be rationalized in terms of the need to protect hydrophobic membrane proteins from aggregation as they pass from ribosomes to mitochondria — a hazardous journey for which such proteins need the

continued protection of chaperones. Cytosolic Hsp70 was known to chaperone the folding of many newly synthesized proteins⁷, whether they are destined for mitochondria or remain in the cytosol, but Hsp90 was thought to be largely restricted to the folding and assembly of signal-transduction proteins in the cytosol⁸. Why mammalian mitochondria require Hsp90 for carrier-protein import whereas yeast mitochondria do not is just one of the questions raised by these observations¹. The study of molecular chaperones continues to spring surprises. ■

R. John Ellis is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

e-mail: jellis@bio.warwick.ac.uk

1. Young, J. C., Hoogenraad, N. J. & Hartl, F. U. *Cell* **112**, 41–50 (2003).
2. Deshaies, R. J. *et al.* *Nature* **332**, 800–805 (1988).
3. Murakami, H., Pain, D. & Blobel, G. *J. Cell Biol.* **107**, 2051–2057 (1988).
4. Ellis, R. J. *Biochem. Biophys. Res. Commun.* **238**, 687–692 (1997).
5. Pfanner, N. & Geissler, A. *Nature Rev. Mol. Cell Biol.* **2**, 339–349 (2001).
6. Brix, J., Dietmeier, K. & Pfanner, N. *J. Biol. Chem.* **272**, 20730–20735 (1997).
7. Frydman, J. *Annu. Rev. Biochem.* **70**, 603–649 (2001).
8. Young, J. C., Moarefi, I. & Hartl, F. U. *J. Cell Biol.* **154**, 267–273 (2001).

Applied physics

Son et lumière

John M. Worlock and Michael L. Roukes

The confinement of photons in a resonant cavity is the basis of laser operation. A device that has a resonant cavity for acoustic phonons inside an optical cavity enhances the interaction between sound and light.

A quintet of physicists has created some interesting 'tunes' using coupled resonant cavities. The double-resonator structures described by Trigo and colleagues¹ in *Physical Review Letters* are unusual in that they are designed to resonate at optical frequencies (with photons) and at acoustical frequencies (with phonons)—simultaneously.

Resonant cavities are ubiquitous. They form an essential part, for example, of orchestral instruments such as brass and woodwind, of pipe organs, and of the vocal cavities of many species. Their study also has a long and illustrious history. In 1877, Lord Rayleigh published his *Theory of Sound*, launching, among other things, the study of waves in confining geometries, or resonators. A notable early optical resonator is that of Fabry and Perot, unveiled in 1897, which has served the optical community well over the intervening years, and whose variants are crucial elements in the feedback systems of lasers. The function of a resonator is to provide space for wave propagation within reflective walls, so as to support a standing wave with a particular wavelength, and hence frequency. But to be useful, the

resonator walls must not be perfectly reflective — they must radiate some energy to the outside world.

Now Trigo *et al.* have performed the first studies of the scattering of standing-wave photons from standing-wave phonons by placing an acoustic cavity inside an optical cavity, carefully positioning the acoustic cavity at the peak of the optical amplitude to maximize the strength of the interaction between light and sound. Their approach may engender new possibilities for the generation of coherent (single-wavelength) phonons, and may even ultimately help to enhance coherence in quantum electronic devices of the future, through the modification of electron–phonon interactions.

The cavities used by Trigo *et al.* were grown by molecular beam epitaxy — the well-developed technique of atomic layer-by-layer growth — as a semiconducting superlattice with alternating layers of AlAs and a related alloy, AlGaAs (Fig. 1). To understand how these structures behave, it is helpful to start with the notion of a Bragg mirror, a periodic assemblage of identical planes that is highly reflective only for certain bands of

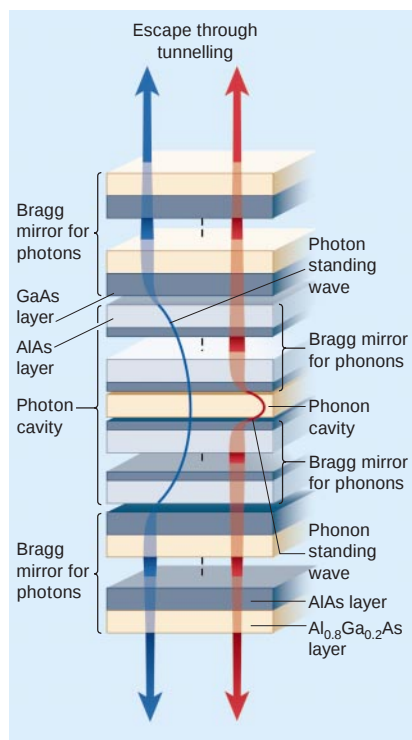


Figure 1 A cavity within a cavity. Trigo *et al.*¹ arranged layers of semiconducting material to make Bragg mirrors (which reflect certain frequencies but transmit others) for photons and phonons, creating an acoustic cavity inside an optical cavity. When a light beam of an appropriate wavelength enters the structure, it sets up a standing wave in the cavity. The light-sound interaction allows a photon of the incident beam to split into two parts: a phonon resonant in its cavity and a photon resonant, and trapped, in its cavity. But as the mirrors are not perfectly reflective, both light and sound can escape the cavity.

photon frequencies, in much the same way that crystal planes reflect X-rays and that electronic forbidden-energy bands occur in crystals. In such a one-dimensional optical Bragg structure, introducing an additional phase-shifting layer that is different from the others can create an optical bound state whose distinct resonant frequency lies within the range that is reflected at the Bragg mirror. The structure then provides an efficient transmission channel: light of an appropriate frequency passes through the Bragg lattice, then is phase-shifted and resonantly enhanced within what is, in essence, a 'microcavity' in the additional layer; gradually the light escapes the microcavity through tunnelling and passes out of the Bragg lattice at its original frequency.

Various authors^{2,3} (in fact, combinations and subsets of the present authors) have made good use of such microcavities in a range of experiments, using optical resonances to enhance processes such as inelastic or Raman scattering. The new twist in this

experiment by Trigo *et al.*¹ is the insertion of a second microcavity within the optical one.

The second cavity is conceived in direct analogy to the first, and is constructed as a resonator for acoustic phonons. Here, the Bragg mirrors have periods of about 10 nm, the central cavity is only 5 nm thick, and the whole structure fits snugly inside the central layer of the optical cavity (Fig. 1). The photons, with wavelengths of hundreds of nanometres, are unaffected by the thin acoustic layers (they simply take the appropriate average refractive index). The resonant frequency of the acoustic phonons in this structure is about 5×10^{11} Hz, well beyond the range of frequencies studied by either microwave or normal Brillouin scattering techniques.

In Trigo and colleagues' experiment, an incident photon is 'scattered' into the resonant, localized state in the optical microcavity, and at the same time a phonon is generated in the resonant, localized state in the acoustic cavity. The incident or absorbed photon must carry energy equal to the sum of the energies of these states and, to achieve resonance and the highest possible efficiency, the photon must be incident at a small angle to the structure (not quite 'head-on'). To resume its journey, and be observed, the 'scattered' photon must escape from the cavity by tunnelling through the Bragg mirror. As the resonant states inside the cavity are standing waves, there is no distinction between forward and backward scattering, and the scattered photon tunnels out simultaneously in both directions. The spectrum of scattered light measured by Trigo *et al.* clearly shows the peak expected for such a device.

Returning to the resonant, or localized, phonon that is produced alongside the scattered photon, it also tunnels out through the Bragg mirrors that form the acoustic cavity. This raises the prospect of realizing a coherent monochromatic source of very-high-frequency acoustic phonons. Trigo *et al.* have not observed these phonons directly, and their intensity outside the microcavity must be very small. But it is not difficult to imagine the steps towards enhancing the phonon intensity — by raising the intensity of the incident light, and also by seeding the optical cavity with photons to stimulate the scattering process. ■

John M. Worlock is in the Department of Physics, University of Utah, Salt Lake City, Utah 84112, USA.

e-mail: worlock@nova.physics.utah.edu

Michael L. Roukes is in the Condensed Matter Physics Group, California Institute of Technology, Pasadena, California 91125, USA.

e-mail: roukes@caltech.edu

1. Trigo, M., Bruchhausen, A., Fainstein, A., Jusserand, B. & Thierry-Mieg, V. *Phys. Rev. Lett.* **89**, 227402 (2002).
2. Fainstein, A., Jusserand, B. & Thierry-Mieg, V. *Phys. Rev. Lett.* **75**, 3764–3767 (1995).
3. Fainstein, A. *et al.* *Phys. Rev. Lett.* **86**, 3411–3414 (2001).



100 YEARS AGO

The fact that the message from the King to President Roosevelt, in reply to the latter's wireless telegram greeting, had to be sent to America by cable occasioned at the time much comment and correspondence in the daily papers on the attitude of the Post Office towards Mr. Marconi... Mr. Marconi made the following statements:—"We asked the Post Office authorities whether they would allow us to connect our station at Poldhu by wire with Mullion — at our own expense, mind you — but they refused absolutely and entirely. The message (that from the King) was not received at our offices until after Mullion Post Office had closed for the night, and one cannot very well keep a King's message lying about for twelve hours. I think it would have been much more discourteous to the King to have kept his message waiting for a day than it was to send it by cable."... In these circumstances, it is not surprising that Mr. Marconi's feelings towards the Post Office are rather bitter... He now proposes to go to Italy and build a huge station there, partly, no doubt, because, as he says, "Abroad I can get everything I want. Here in England I can get nothing." This is a little sweeping, for all England has not been so backward in supporting Mr. Marconi's enterprise as the officials of the Post Office. From *Nature* 19 February 1903.

50 YEARS AGO

We have formulated a structure for the nucleic acids which is compatible with the main features of the X-ray diagram and with the general principles of molecular structure, and which accounts satisfactorily for some of the chemical properties of the substances. The structure involves three intertwined helical polynucleotide chains. Each chain, which is formed by phosphate di-ester groups and linking β -D-ribofuranose or β -D-deoxyribofuranose residues with 3', 5' linkages, has approximately twenty-four nucleotide residues in seven turns of the helix. The helices have the sense of a right-handed screw. The phosphate groups are closely packed about the axis of the molecule, with the pentose residues surrounding them, and the purine and pyrimidine groups projecting radially, their planes being approximately perpendicular to the molecular axis. The operation that converts one residue to the next residue in the polynucleotide chain is rotation by about 105° and translation by 3.4 Å. Linus Pauling, Robert B. Corey. From *Nature* 21 February 1953.

Taste

Flavour of the week in signalling pathways

Cell **112**, 293–301 (2003)

Tastes come in five flavours: salty, sour, bitter, sweet or umami — the taste of amino acids. Taste buds in the mouth contain many taste-receptor cells, which relay flavours, via nerves, to taste centres in the brain. According to one theory, a single taste-receptor cell carries receptors for the different flavours, each using a different intracellular signalling pathway to convey the sensation to nerves. But Yifeng Zhang *et al.* find that sweet, bitter and umami are detected by distinct cells, using different receptors but the same signalling pathway.

The sweet, bitter and umami receptors known are mainly found on separate taste-receptor cells. But all these cells contain the enzyme phospholipase C β 2 and the ion channel TRPM5. Zhang *et al.* show that mice lacking either phospholipase C β 2 or TRPM5 cannot taste sweet, bitter or umami, yet have normal salt and sour perception. The authors propose that phospholipase C β 2 and TRPM5 are part of the same signalling pathway, in which the former activates the latter. Restoring phospholipase C β 2 only in cells carrying bitter receptors restores the response to bitter tastes, but not to sweetness or umami, confirming that individual taste-receptor cells detect only one of these three flavours. **Marie-Thérèse Heemels**

Atomic imaging

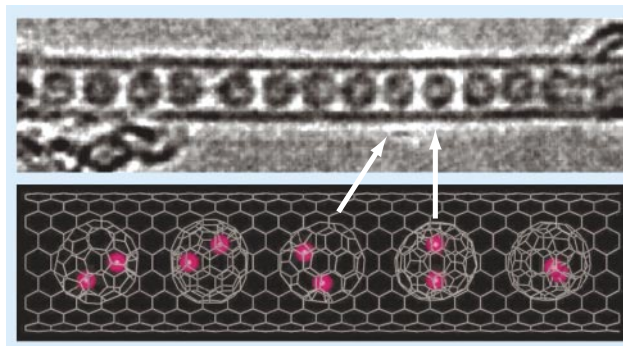
Like peas in a 'nanopod'

Phys. Rev. Lett. **90**, 055506 (2003)

The cage-like structure of carbon fullerenes can encapsulate other elements, and the properties of these complexes depend strongly on the precise location of the guest atoms. This is particularly true of the dimetallofullerenes — complexes that contain two metal atoms within each fullerene molecule.

Using high-resolution transmission electron microscopy (HRTEM), K. Suenaga and colleagues have determined the location of metal atoms in individual molecules of scandium dimetallofullerene, Sc $_2$ @C $_{84}$. The scandium atoms undergo both covalent and ionic interactions with their C $_{84}$ cage, and three different isomers of this molecule have been predicted. But one isomer, with a peculiar nonpolar arrangement of scandium atoms, had never been observed.

The use of HRTEM to directly identify specific atomic species, particularly such a light element as scandium, is challenging. Suenaga *et al.* filled a single-walled carbon nanotube with Sc $_2$ @C $_{84}$ molecules purified



In this atomic-resolution image (top) of dimetallofullerene molecules inside a carbon nanotube, the positions of the scandium atoms encapsulated in the fullerene can be seen to match the theoretical prediction (bottom).

to contain only the isomer of interest. By comparing their HRTEM images of these fullerene 'peapods' with simulations (see picture), the authors identified the positions of the scandium atoms in the fullerene isomer, and confirm that they match the theoretical prediction. **Ed Gerstner**

Animal behaviour

Heredity and the bolder bird

Proc. R. Soc. Lond. B doi:10.1098/rspb.2002.2300 (2003)

Bold great tits have adventurous children, whereas the offspring of timid birds stay closer to home. The discovery of a link between birds' personality and their dispersal behaviour points to this trait as having a powerful effect on the genetic structure of populations.

Niels J. Dingemanse and his colleagues captured wild great tits (*Parus major*) and put them in a novel environment — a sealed room containing artificial trees. Some birds were quick to explore their surroundings, others more hesitant. After the birds' return to the real world, the researchers found that the offspring of the bolder individuals dispersed the farthest, whereas those of timid tits stayed close to the parental nest. They also found that birds that had migrated into the study population were more adventurous than locally born individuals, and that bold females — the sex that, in birds, tends to leave home — dispersed farther after fledging than shy ones.

About half of the variation in a great tit's personality can be explained by heredity. So the geographical distribution of genes for boldness will affect the spread of other genes, and also the population dynamics of the species — what birds colonize new environments, for example. **John Whitfield**

Biophysics

Unzipping your genes

Proc. Natl Acad. Sci. USA **100**, 1694–1699 (2003)

The unzipping of DNA in the cell for replication typically proceeds in stops and starts. Generally, this seems to be caused by faults or twists, which prompt pauses for

DNA repair. But experiments by Claudia Danilowicz *et al.* show that even flawless DNA unzips sporadically. The researchers tethered one strand of viral DNA to a glass capillary, and pulled, with constant force, on a microscopic magnetic bead attached to the other strand. The separation of the strands occurred in a series of jumps, repeatedly becoming stuck for several minutes or so.

This behaviour is expected, because G–C base pairs are slightly more strongly bound together than A–T pairs. So G–C-rich sequences will be harder to prise apart. Because unbinding is thermally activated, the duration of the stalled spells is randomly determined. If the separation force is small enough, the unzipping process may pause for hours on end. **Philip Ball**

Neurobiology

Firing patterns add up

Neuron **37**, 499–511 (2003)

When nerve cells signal from the eye to the brain, they do so in harmony, say Mark J. Schnitzer and Markus Meister. They have built a new algorithm that can detect the cells talking in unison.

Groups of neurons are thought to carry coded information about the world by firing in coordinated sequences. But computers struggle to detect these patterns from the vast number of possible combinations hidden in the thousands of signals recorded from neurons.

To get round this, Schnitzer and Meister built an iterative algorithm. Rather than looking for a firing pattern among all signals and all cells, it first looks for a match between pairs of cells — and then searches for others that fire in sync. Using their program, the researchers found sets of up to seven cells that signal together in the salamander's retina. These account for more than half of all the electrical 'spikes' they recorded. The authors suggest that each firing pattern, rather than each neuron, carries a separate message to the brain. This could communicate better visual detail about an image, because there are many more possible patterns than individual neurons. **Helen Pearson**

Severe summertime flooding in Europe

Even as summers become drier, the incidence of severe precipitation could increase.

Using a high-resolution climate model, we are able to quantify the influence of greenhouse-gas-induced global warming upon heavy or extended precipitation episodes that inflict catastrophic flooding. We find that an increase in the amount of precipitation that exceeds the 95th percentile is very likely in many areas of Europe, despite a possible reduction in average summer precipitation over a substantial part of the continent. Our results indicate that episodes of severe flooding may become more frequent, despite a general trend towards drier summer conditions.

In the European Union project PRUDENCE (EVK2-CT-2001-00132; ref. 1), the high-resolution (50-km grid) regional climate model HIRHAM4 (ref. 2) created by the Danish Meteorological Institute has been applied to two of the emission scenarios, A2 and B2, drawn up by the Intergovernmental Panel on Climate Change (IPCC)³. Three 30-year time-slice experiments were carried out for periods representing roughly the present (1961–90) and two future scenarios (2071–2100), respectively. The large-scale controlling conditions originated from transient climate-change simulations using the coupled ocean–atmosphere global climate model (OAGCM) ECHAM4/OPYC (300-km grid; refs 4, 5).

Several studies with OAGCMs have indicated that precipitation in many regions of the Earth could change considerably in a warmer climate⁶. Simulations consistently predict that summer precipitation will be reduced in many mid-latitude regions, whereas at higher latitudes there will be little change or possibly an increase⁶. Although OAGCMs indicate an intensified global hydrological cycle in a warmer climate, evidence of changes in extreme precipitation at the regional scale remains unconvincing⁷. The low spatial resolution of OAGCMs precludes a realistic simulation of regional circulation and therefore of extreme precipitation.

We investigated the relationship between climate change and heavy precipitation episodes lasting for 1–5 days at high horizontal resolution during the late boreal summer (July–September). During this period, European climatic conditions occasionally favour very severe precipitation episodes, such as those that caused the recent flooding of the rivers Odra (1997), Elbe (2002) and of sub-catchments of the Rhone (2002).

Figure 1a shows the relative change in mean precipitation during July, August and

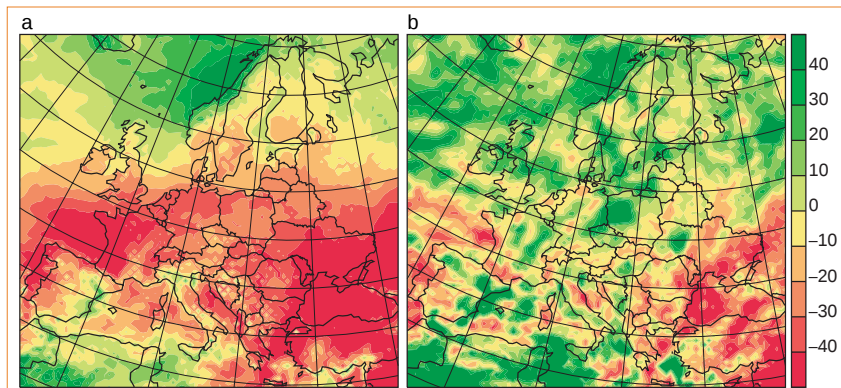


Figure 1 Relative percentage change in precipitation for July–September in the Intergovernmental Panel on Climate Change's A2 scenario with respect to the present day. Relative change is shown for **a**, the seasonal mean; and **b**, the five-day mean exceeding the 99th percentile.

September according to HIRHAM4 for the A2 scenario (the warmer of the two) with respect to present conditions. A two-sided *t*-test shows that all changes exceeding ± 5 –10% of the present-day mean are significant at the 95% level. The large-scale patterns of change are similar to those in the driving OAGCM model.

Model validation of extremes is generally limited by insufficient observational data: European daily-precipitation data, in particular, are not easily available. In two representative cases (Scandinavia and the Alps), however, the HIRHAM4 model can simulate relatively realistic precipitation frequencies, although it has a tendency to underrepresent the most extreme events^{1,8,9}.

The relative change in the mean five-day precipitation for July–September that exceeds the 99th percentiles in scenario A2 with respect to the control is shown in Fig. 1b. Even when a reduction in total mean precipitation is simulated, the amounts of precipitation in the intensive events are much less reduced, and even increase in many places. The higher the percentile considered, the larger are the areas that show a positive change. A standard Mann–Whitney test on a series of events that exceed different thresholds indicates that there is a significant change at the 95% level for events above the 95th percentile, but the small sample size precludes a general, unequivocal detection of change for the 99th percentile (or higher).

This, however, does not necessarily imply the absence of change. These changes are seen for a wide range of percentiles and for most European land points, with exceptions being mostly located on the Iberian Peninsula and over the Balkans. To illustrate the shift clearly while retaining a meaningful sample size, we show the 99th

percentile in Fig. 1b, even though only those changes that exceed about $\pm 10\%$ are statistically significant.

We also carried out another similar analysis, averaging the precipitation data over several European river catchments. With few exceptions (involving rivers situated in a very dry environment), an increase in the amount of precipitation during extreme precipitation episodes is simulated in both scenarios (most prominent in A2).

Using the combined information from both simulations, we find that CO₂-induced warming can lead to a shift towards heavier intensive summertime precipitation over large parts of Europe; the warmer scenario (A2) gives the largest shift. This finding may be explained¹⁰ by the fact that the atmosphere will contain more water in a warmer climate (according to the Clausius–Clapeyron equation), which will provide further potential for latent-heat release during the build-up of low-pressure systems, thereby possibly both intensifying the systems and making more water available for precipitation.

Jens H. Christensen, Ole B. Christensen
Danish Meteorological Institute, Lyngbyvej 100,
2100 Copenhagen Ø, Denmark
e-mail: jhc@dmi.dk

- Christensen, J. H., Carter, T. R. & Giorgi, F. *EOS* **83**, 147 (2002).
- Christensen, O. B., Christensen, J. H., Machenhauer, B. & Botzet, M. J. *Clim.* **11**, 3204–3229 (1998).
- Nakićenović, N. et al. *IPCC Special Report on Emission Scenarios* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Roegner, E. et al. *MPI Report 218* (Max Planck Inst., Hamburg, 1996).
- Roegner, E., Bengtsson, L., Feichter, J., Lelieveld, J. & Rohde, H. J. *Clim.* **12**, 3004–3032 (1999).
- Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, UK, 2001).
- Voss, R., May, W. & Roegner, E. *Int. J. Clim.* **22**, 755–777 (2002).

8. Frei, C. et al. *J. Geophys. Res.* (in the press).
9. Christensen, O. B., Christensen, J. H. & Botzet, M. in *Implications of Climatic Change for the Hydrological Cycle and for Water Management* (ed. Beniston, M.) 101–112 (Kluwer Academic, Dordrecht, The Netherlands, 2001).
10. Frei, C., Schär, C., Lüthi, D. & Davies, H. C. *Geophys. Res. Lett.* **25**, 1431–1434 (1998).

Competing financial interests: declared none.

Ecology

Parthenogenesis in an outsider crayfish

It has been rumoured¹ that an unidentified decapod crustacean, a crayfish of marbled appearance and of uncertain geographical origin that was introduced into the German aquarium trade in the mid-1990s, is capable of unisexual reproduction (parthenogenesis). Here we confirm that this marbled crayfish (*'Marmorkrebs'*) is parthenogenetic under laboratory conditions and use morphological and molecular analysis to show that it belongs to the American Cambaridae family. Although parthenogenesis is widespread among the Crustacea², and shrimp, lobsters, crayfish and crabs are otherwise versatile in their modes of reproduction^{3–5}, it has not been reported before in decapods, the largest and economically most important

crustacean group. By virtue of its parthenogenetic reproduction, the marbled crayfish emerges not only as an interesting laboratory model but also as a potential ecological threat in that it could outcompete native forms should even a single specimen be released into European lakes and rivers.

To determine the origin and phylogenetic position of the marbled crayfish (Fig. 1a), we compared the sequences of sections of two mitochondrial genes with those of other related species, with particular reference to a similar-looking cambarid species, *Procambarus fallax*, from Florida. The sequences in both genes of the marbled crayfish differ at only a few positions from those of other cambarids, which supports their morphological similarity (the presence of a spermatheca)⁶. Our phylogenetic analysis indicates a particularly close affinity with *P. fallax*, although the marbled crayfish's species identity remains to be verified (Fig. 1b).

We tested whether *Marmorkrebs* could be parthenogenetic by studying a mature female from a laboratory population in Berlin whose spermatheca contained no evidence of spermatophores from copulation, but which repeatedly laid eggs. Between two broods, this crayfish moulted, a procedure that clears any remnants of sperm from the spermatheca. We also sexed laboratory populations in Berlin (93 specimens from 7 mothers of two generations) and Heidelberg (39 specimens), beginning at the earliest stage at which sex can be determined. We found that all specimens (body length, 0.8–8.0 cm) showed female morphology, which excludes protandric hermaphroditism as the mode of reproduction. No spermatophore was detected in our study.

To rule out internal autogamy, we studied the histology and ultrastructure of the reproductive system of all 39 specimens of the Heidelberg population from juvenile (1.9 cm) to post-brooding (6.8 cm) stages. All gonads were normal ovaries with oviducts, and there was no evidence of ovotestes or male gonoducts (results not shown), which are usually present in hermaphroditic crayfish⁷. Azan staining of some adult gonad sections revealed an abundance of large, primary vitellogenic oocytes, and proliferating clusters of small pre-vitellogenic oocytes in the presence of hatchlings, indicating that a new reproductive cycle was already under way (Fig. 1c). These results provide convincing evidence for parthenogenesis in the marbled crayfish.

Our findings have several practical implications. This marbled crayfish will be

useful in the laboratory for physiological, ecological, evolutionary, developmental and genetic studies, having the advantages of fitness, high fertility, fast growth, unisexuality and isogenic progeny⁸. The large oocytes are easily accessible for genetic manipulation, making this species a candidate model for transgenesis⁹ in decapod crustaceans. The rapid reproduction of this crayfish might also be of interest for commercial farming purposes.

Last but not least, this crayfish, which is now widespread in Europe's aquaria, could become a menace to European freshwater ecosystems, as the release of even one specimen into the wild would be enough to found a population that might outcompete native crayfish. As an American species, it is a potential transmitter of the infectious crayfish plague that almost caused the extinction of the native European crayfish and which still threatens wild and farmed populations¹⁰.

Gerhard Scholtz*, Anke Braband*, Laura Tolley†, André Reimann*, Beate Mittmann*, Chris Lukhaup‡, Frank Steuerwald†, Günter Vogt‡

*Humboldt-Universität zu Berlin, Institut für Biologie/Vergleichende Zoologie, 10115 Berlin, Germany

e-mail: gerhard.scholtz@rz.hu-berlin.de

†Universität Heidelberg, Institut für Zoologie, INF 230, 69120 Heidelberg, Germany

‡Gotenstrasse 16, 1336 Bittenfeld, Germany

1. Lukhaup, C. *Aquaristik Aktuell* **7–8**, 48–51 (2001).
2. Gruner, H.-E. in *Lehrbuch der Speziellen Zoologie, Band I, 4. Teil: Arthropoda* (ed. Gruner, H.-E.) 448–1030 (Fischer, Jena, Germany, 1993).
3. Bauer, R. T. & Holt, G. *J. Mar. Biol.* **132**, 223–235 (1998).
4. Duffy, J. E. *Nature* **381**, 512–514 (1996).
5. Reynolds, J. D. in *The Biology of Freshwater Crayfish* (ed. Holdich, D. M.) 152–191 (Blackwell Science, Oxford, 2002).
6. Scholtz, G. in *The Biology of Freshwater Crayfish* (ed. Holdich, D. M.) 30–52 (Blackwell Science, Oxford, 2002).
7. Vogt, G. in *The Biology of Freshwater Crayfish* (ed. Holdich, D. M.) 53–151 (Blackwell Science, Oxford, 2002).
8. Vrijenhoek, R. C. *Annu. Rev. Ecol. Syst.* **25**, 71–96 (1994).
9. Nam, Y. K., Cho, Y. S. & Kim, D. S. *Transgenic Res.* **9**, 463–469 (2000).
10. Söderhäll, K. & Cerenius, L. *Freshwat. Crayfish* **12**, 11–35 (1999).

Competing financial interests: declared none.

correction

Ultrasound polycrystalline diamond from graphite

T. Irifune, A. Kurio, S. Sakamoto, T. Inoue, H. Sumiya *Nature* **421**, 599–600 (2003)

In the legend to Fig. 1a of this communication, the diameter of the transparent polycrystalline diamond shown is 1 mm, and not 0.1 mm as published; the scale divisions represent 0.1 mm. Also, the first full paragraph in the second column on page 600 should read: "Recent chemical-vapour deposition techniques provided pure polycrystalline diamonds, but these diamonds are not sintered and have poor intergrain adhesion. Accordingly, they have been reported to have a hardness of ~80–100 GPa (ref. 10), which is significantly lower than the highest value (~120 GPa; ref. 4) for single-crystal diamonds."

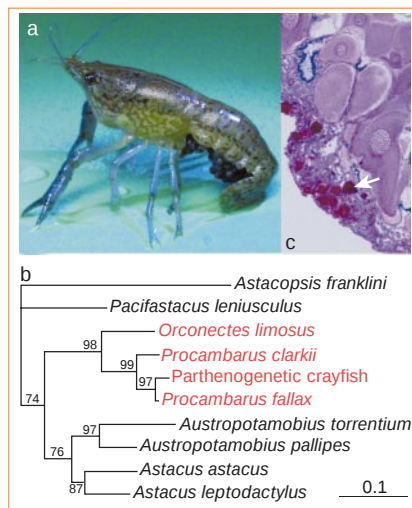


Figure 1 Characteristics of the marbled crayfish. **a**, Lateral view of the crayfish, showing the eggs attached to its pleopods. Adult specimens are 3–10 cm in length (rostrum to telson). The basic colour is brown–green, with light and dark spots. **b**, Phylogenetic relationships of the marbled crayfish based on a comparison of the partial sequences of two mitochondrial genes, *Cox1* and the gene encoding 12S ribosomal RNA; GenBank accession numbers, AY151515–AY151524, AY151525–AY151534. Maximum-likelihood analysis used the Hasegawa–Kishino–Yano model, assuming rate heterogeneity with a log likelihood of -3075.69 ; the reliability of branches is given as a percentage of puzzling steps where the appropriate node arises. Parsimony analysis produced similar results. The marbled crayfish belongs to the American Cambaridae (red) and is closely related to *Procambarus fallax* (only 2.2% of base positions different). Scale bar, 0.1 nucleotide substitutions. **c**, Azan-stained histological section of a gonad of a female with hatchlings exclusively composed of ovarian tissue. Arrow, pre-vitellogenic oocyte; arrowhead, primary vitellogenic oocyte.

An exceptionally preserved Lower Cretaceous ecosystem

Zhonghe Zhou*, Paul M. Barrett†‡ & Jason Hilton§

* Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China

† Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK

§ Department of Geology and Zoology, National Museums of Scotland, Chambers Street, Edinburgh, EH1 1JF, UK

Fieldwork in the Early Cretaceous Jehol Group, northeastern China has revealed a plethora of extraordinarily well-preserved fossils that are shaping some of the most contentious debates in palaeontology and evolutionary biology. These discoveries include feathered theropod dinosaurs and early birds, which provide additional, indisputable support for the dinosaurian ancestry of birds, and much new evidence on the evolution of feathers and flight. Specimens of putative basal angiosperms and primitive mammals are clarifying details of the early radiations of these major clades. Detailed soft-tissue preservation of the organisms from the Jehol Biota is providing palaeobiological insights that would not normally be accessible from the fossil record.

The Yixian and Jiufotang Formations of northeastern China have yielded a huge variety of exceptionally well-preserved fossils that comprise the 'Jehol Biota'^{1–8}. Recent discoveries of spectacular vertebrate specimens, particularly of theropod dinosaurs, birds and mammals with associated integumentary structures and other soft tissues, and of putative early angiosperms have established the global significance of this Lower Cretaceous biome^{9–16}. The Jehol Biota includes representatives of almost all of the major clades of Lower Cretaceous terrestrial and freshwater vertebrates, a wide variety of invertebrates and a diverse flora. It is unique in terms of its species richness, exceptional preservation and in the sheer numbers of specimens of many of the species present. The combination of these factors means that the Jehol Biota provides a rare, incredibly detailed picture of an intact Early Cretaceous terrestrial ecosystem. Here, we present a synthetic overview of the entire biota, emphasising its impact on wide-ranging evolutionary, palaeoecological and phylogenetic questions, including the origin of birds, the evolution of feathers and flight, the early diversification of angiosperms, and the timing of the placental mammal radiation.

Geological, palaeoenvironmental and preservational setting

The Jehol Group comprises the Yixian and Jiufotang Formations, which crop out in western Liaoning, northern Hebei and south-eastern Inner Mongolia (Figs 1 and 2). Lateral equivalents of these units, with similar biotas, are found in adjacent areas of eastern and central Asia, including Kazakhstan, Mongolia, Siberia, Japan and Korea (Fig. 1)^{3–4}. The Yixian and Jiufotang Formations are conformable, lithologically similar deposits of weakly laminated to finely bedded siliciclastic sediments, mainly low-energy sandstones and shales, intercalated with extrusive basalts and tuffs and cross-cut by occasional dykes and sills¹⁷ (Fig. 2). Jehol Group sediments represent freshwater lacustrine environments, and lack the laterally variable features of other freshwater settings such as rivers and deltas¹⁷, a reconstruction that is consistent with the mixed terrestrial/freshwater fossil assemblage. Eastern Asia was a fully emergent landmass during the late Mesozoic era¹⁸, and this terrestrial depositional setting contrasts with those of most other penecontemporaneous biotas, which show evidence of greater marine influence or are exclusively marine. For example, the sediments of the Santana and Crato Formations of Brazil (Aptian/Albian) represent saline

lagoons, whereas the Spanish locality of Montsec (Berriasian/Valanginian) was deposited under peritidal conditions¹⁹. Extensive volcanism, resulting from increased tectonic activity along the Pacific Rim at this time²⁰, is apparent from the conformable deposition of tuffaceous sediments within the Jehol Group. Volcanic activity was most frequent during deposition of the Yixian Formation, but became progressively less pronounced in Jiufotang times¹⁷. Regional volcanism, in combination with the presence of numerous shallow lakes, provided ideal environments for the exceptional preservation of the Jehol Biota and has also permitted high-resolution dating of the fossiliferous horizons (see below). Palaeobotanical and sedimentological evidence indicate seasonal climatic fluctuations between semi-arid and mesic conditions^{4,21}.

Freshwater and terrestrial organisms from the biota usually occur together within the same sedimentary horizon (Fig. 2). Preservation of complete articulated shells, arthropod exoskeletons, vertebrate skeletons and terrestrial plant stems with associated leaves, rootlets and other structures, indicates that all of these specimens originated in close proximity to low-energy lacustrine depositional sites and were not transported over extensive distances. Individual elements are generally unbroken and display little or no abrasion^{6,22}.

Dead organisms entered the lakes, were buried rapidly and encased in fine-grained sediments. The most productive horizons are beneath ash tuff falls, which would have entombed most of the organisms present in the water column—the tuff layers are strongly correlated with mass mortality events^{6,22,23}. Tuff deposition effectively sealed the fossil-bearing sediments, creating microenvironments around the organisms that promoted anoxic conditions and halted bacterial decay of soft tissues. Moreover, tuffs prevented burrowing organisms from disturbing organic remains within the sediment, eliminating bioturbation and scavenging. The combination of these factors resulted in the formation of *Konservat-Lagerstätten* conditions¹⁹, permitting exceptional preservation of many original soft-tissue features. Integumentary structures (filaments, feathers and fur) are known from pterosaur²⁴, dinosaur^{13,14,25–30}, bird^{31,32} and mammal³³ specimens: preservation of other soft tissues (skin impressions, cartilaginous elements, keratinous beaks) and stomach contents is also common and has been reported for a wide variety of taxa⁷. Preservation of wing membranes, hairs and colour patterning have been recorded from insects and other invertebrates^{4,7}. The gross morphology of Jehol plants is often well preserved, but fine anatomical detail and cuticular material are usually lacking. Consequently, anatomical features that are generally not preserved in fossils are often found in the Jehol material. Three-dimensionally

‡ Present address: Department of Palaeontology, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK.

preserved specimens of dinosaurs^{34–37}, mammals^{38,39}, frogs and lizards have been recovered from the Lujiatun locality in the Yixian Formation near Beipiao, western Liaoning Province. The fossil-bearing tuffs at this locality lack obvious bedding planes, suggesting that this deposit resulted from a single, catastrophic mass mortality event.

Age of the deposits

The dating of the Yixian and Jiufotang Formations has proved to be contentious^{4,5,8,40–42}, and many biostratigraphical correlations and radiometric dates have been published that support the Jehol Biota being either Late Jurassic or Early Cretaceous in age^{5,8,42}. Establishing the age of these units is critical for evaluating the evolutionary significance of the biota, as the Late Jurassic/Early Cretaceous interval is an important time in Earth history, during which major changes in the composition and dynamics of terrestrial ecosystems occurred⁵. For example, if a Late Jurassic age were accepted, this would suggest that the phylogenetic and ecological radiations of birds, placental mammals and angiosperms began significantly earlier than current palaeontological evidence indicates.

Most of the evidence supporting a Late Jurassic age for the Jehol Group is either equivocal or flawed^{5,8,42,43}. Many proposed Late

Jurassic biostratigraphical correlations are based on taxa with low stratigraphical resolution, use taxa that are difficult to diagnose or differentiate from each other, or depend on vertebrates of limited biostratigraphical utility^{5,8}. ⁴⁰K–⁴⁰Ar and ⁸⁷Rb–⁸⁷Sr dates of 137 ± 7 Myr ago and 143 ± 4 Myr ago, from the Yixian Formation, have been advanced in favour of a Late Jurassic age⁴⁴. However, this conclusion depends on the adoption of a chronological framework in which the Jurassic/Cretaceous boundary occurred at about 135 Myr: in contrast, most geochronologists regard the latter as unreliable, accepting a slightly older age (approximately 144 Myr) for the boundary^{5,8}. Consequently, these ⁴⁰K–⁴⁰Ar and ⁸⁷Rb–⁸⁷Sr dates actually support an Early Cretaceous age when the older boundary age is applied^{5,8,42}. ⁴⁰Ar–³⁹Ar dating of biotite crystals from a tuff in the Yixian Formation suggested a Late Jurassic age, producing an overall mean age of 145.3 ± 4.4 Myr and a combined isochron date of 147.1 ± 0.18 Myr⁴⁵. However, these results are subject to doubt as some evidence suggests that the samples used were either altered diagenetically or contained trapped argon, either of which could adversely affect the results of the analyses⁴².

Current evidence indicates that the Jehol Biota is of late Early Cretaceous age^{5,8,41–43,46}. ⁴⁰Ar–³⁹Ar dates of 124.6 ± 0.1 Myr and 125.0 ± 0.18 Myr obtained from total fusion and incremental heating analyses of sanidine and biotite crystals from three different tuff layers in the 'Jianshangou beds' of the Yixian Formation indicate referral to the Barremian stage^{41,42}, whereas a ⁴⁰Ar–³⁹Ar date of 128.4 ± 0.2 Myr from a basalt capping the lowermost 'Lujiatun beds' suggests a Hauterivian age for the base of the formation⁴⁶. A ²³⁵U–²⁰⁷Pb zircon date of 125.2 ± 0.9 Myr for the Jianshangou beds is in close agreement⁴⁶, and a mean age of 121.1 ± 0.2 Myr obtained from overlying lava layers and intrusive volcanics⁴⁷ adds support to this conclusion^{41,42}. Moreover, ⁴⁰Ar–³⁹Ar dates of 139.4 ± 0.19 Myr obtained from the underlying Tuchengzi Formation (demonstrating an earliest Cretaceous, Berriasian age for this unit)⁴² confirm that the succeeding Jehol Group was deposited during the Cretaceous period. An Early Cretaceous age is also supported by numerous biostratigraphical correlations^{5,8,43}, although some of these present minor conflicts with radiometric dating, placing the deposits in the basalmost Cretaceous (Berriasian/Valanginian stages). An ⁴⁰Ar–³⁹Ar date of 110.59 ± 0.52 Myr has been obtained from an intrusive basalt within the Jiufotang Formation of Inner Mongolia, indicating an Aptian age for this formation in an adjacent region of northeast China⁴⁸. Consequently, we regard the Jehol Biota as Late Hauterivian to Early Aptian in age, and to have existed for a minimum of 18 Myr.

Vertebrate palaeontology

The vertebrate fauna includes osteichthyan fish^{4,49}, lissamphibians (discoglossid anurans⁵⁰, various urodeles⁵¹), chelonians (sinemydids⁴), choristoderes^{52,53}, squamates⁵⁴, pterosaurs (anurognathids^{24,55}, pterodactylids⁵⁶, ctenochasmatis⁵⁵) and dinosaurs (undescribed sauropods²², ornithopods^{36,57}, ceratopians^{30,35,58} and an ankylosaur⁵⁹). Some species (birds, choristoderes, salamanders and fishes) are known from hundreds of specimens. The most significant discoveries are undoubtedly the non-avian coelurosaurian theropods, the diverse avifauna and a variety of mammals, all of which have impacted on wide-ranging evolutionary debates.

Material from the Jehol Biota has greatly augmented and strengthened the hypothesis that birds are direct descendants of theropod dinosaurs, and has provided new insights into the origin of feathers and of flight, as well as the physiological status of dinosaurs and early birds^{60,61}. Its diverse theropod fauna spans the phylogenetic transition from basal coelurosaurians to ornithurine birds (Figs 3 and 4). The basal coelurosaur *Sinosauropteryx* possesses an extensive covering of filamentous integumentary structures¹³ (Fig. 4), which are considered to be the precursors of true feathers²⁷. Identical structures are also known in the therizinosauroid *Beipiao-*

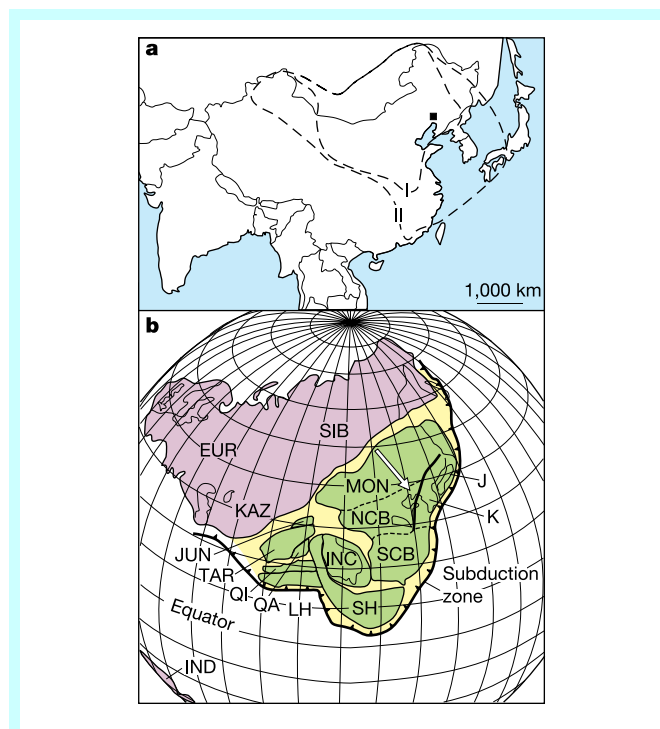


Figure 1 Palaeogeographical setting of the Jehol Biota. **a**, Modern-day map showing different geographical extents of the Jehol Biota through time. I, Yixian Formation and its lateral equivalents (Late Hauterivian/Barremian); II, Jiufotang Formation and its lateral equivalents (Early Aptian); modified from ref. 4. The filled square marks the position of the major vertebrate-bearing sites in Liaoning Province. Insect faunas from penecontemporaneous units in Kazakhstan are remarkably similar to those from the Jehol Biota, and may indicate a westward extension of these ranges⁷¹.

b, Palaeogeographic map of eastern Asia in the Lower Cretaceous, showing major regional tectonic features (modified from ref. 99). The arrow indicates the approximate position of outcrop of the Yixian and Jiufotang Formations in northeastern China. This region would have occupied a palaeolatitude of approximately 40–45° N during the late Mesozoic era^{18,99}. Abbreviations refer to major tectonic divisions: EUR, Europe; INC, Indo-China; IND, India; J, Japan; JUN, Junggar; K, Korea; KAZ, Kazakhstan; LH, Lhasa; MON, Mongolian; NCB, north China; QI, Qiangtang; SCB, south China; SH, Shan Thai; SIB, Siberian; TAR, Tarim.

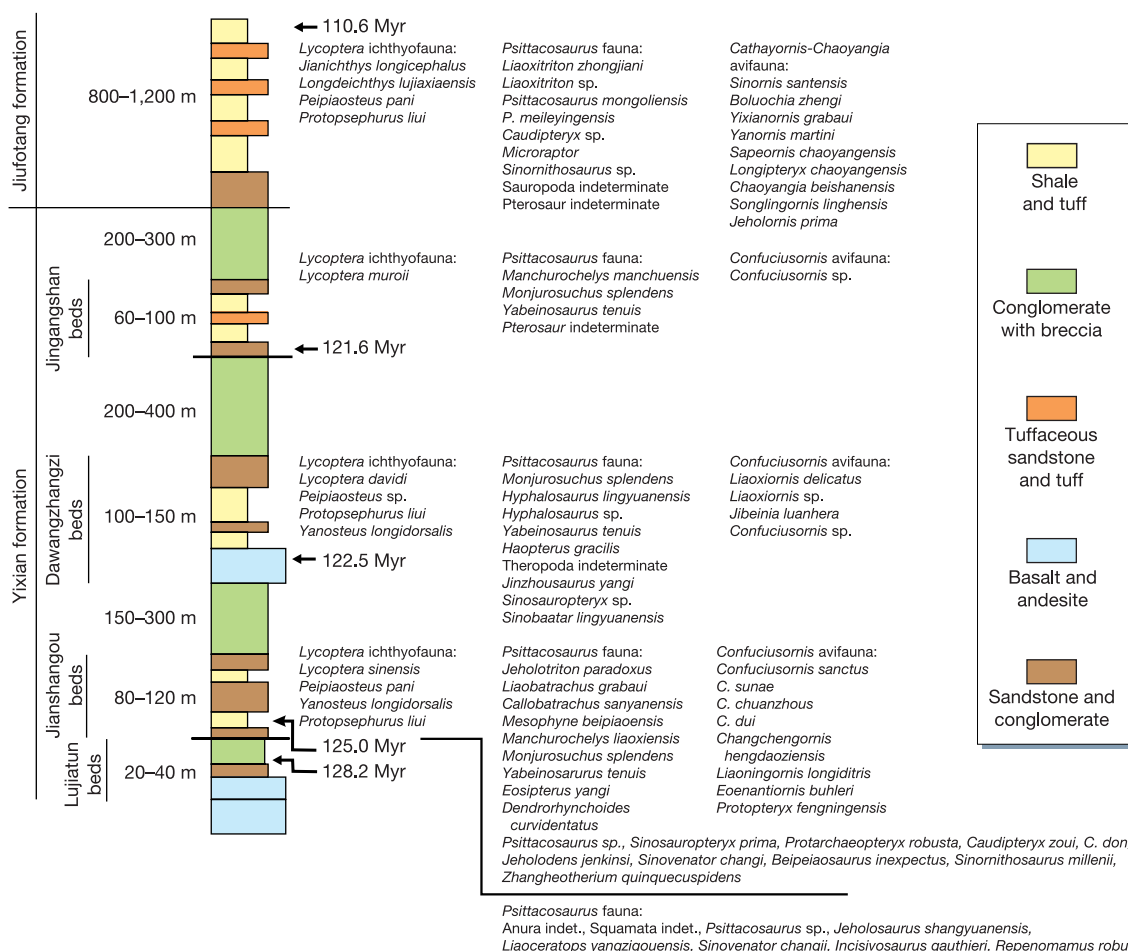
*saurus*²⁵. In contrast, the basal oviraptorosaurian *Caudipteryx* and the enigmatic coelurosaur *Protarchaeopteryx* definitely bear feathers of modern appearance¹⁴, in which a central rachis supports a branching pattern of barbs (Fig. 4). Similar feathers are present in Jehol dromaeosaurs, including *Sinornithosaurus*^{26,27} and several undescribed taxa^{28,29}. All of these animals lacked flight specializations, and most were ground-dwelling cursorial animals, indicating that the selective pressures responsible for initiating the development of feathers were imposed before the origin of flight^{60,61}. This conclusion is underlined by the appearance of 'protofeathers' in basal coelurosaurians that were distant relatives of birds^{13,27,61} (Fig. 3). Increasing feather complexity in oviraptorosaurians and dromaeosaurids accords well with their status as progressively more derived avian outgroups (Figs 3 and 4). Recognition of true feathers in these non-avian dinosaurs demonstrates that feathers are not unique to birds, but had a broader phylogenetic distribution within Theropoda^{60,61}.

As the origin of feathers is not correlated with the origin of flight, other functions must be sought for protofeathers and the earliest true feathers⁶¹. Sexual display, camouflage, species recognition and thrust generation during running/jumping are plausible hypotheses that may account for the origin of varied feather types, although all of these suggestions are difficult to test on the basis of palaeontological evidence⁶¹. These integumentary structures would certainly have had a considerable effect on the thermoregulatory strategies of

these animals, and may have had an important role in the evolution of brooding behaviour⁶¹.

The presence of feathers in *Caudipteryx* and *Protarchaeopteryx* has been dismissed as a vestige of secondary flightlessness, and the identification of these structures as integumentary features has been questioned⁶², but these interpretations have not been supported by other studies. Phylogenetic evidence demonstrates emphatically that these animals were primarily flightless non-avian theropods, not secondarily flightless basal birds, and detailed anatomical observations confirm the integumentary derivation of these structures⁶¹. Moreover, the feathers of *Sinornithosaurus* and integumentary structures of *Sinosauropteryx* are consistent with models of feather evolution and development based on neontological evidence²⁷.

Microaptor, another Jehol theropod with true feathers, is the smallest known dromaeosaurid (trunk length approximately 50 mm)⁶³. It exhibits several pedal characteristics that have been suggested as potentially consistent with a climbing/perching mode of life, whereas other features of the skeleton demonstrate that it was not a volant or secondarily flightless animal. The discovery of this possibly arboreal dromaeosaurid has contributed to the debate over the origin of flight, and has been cited as evidence for the now heterodox view that flight originated from the trees down (by gliding and jumping down from trees), rather than from the ground up (by running and leaping for prey)⁶³. However, detailed



functional morphological studies on *Microraptor* have not yet been attempted and are needed to test its proposed arboreal mode of life.

The high diversity and abundance of birds in the Jehol fauna provides a unique window onto the adaptive radiation and palaeoecology of early avians^{64–67}. Body size and locomotory differences between sympatric Jehol taxa indicate that a rapid ecological diversification of avian taxa occurred during the Early Cretaceous^{66,67}. Several taxa (*Jeholornis*, *Sapeornis* and *Confuciusornis*) lie outside Ornithothoraces (Fig. 3) and help to determine the sequence of acquisition of typical avian features, such as cranial kinesis, the horny beak, and the pygostyle (Figs 3 and 4)^{31,64,67}. *Confuciusornis* is represented by hundreds of specimens from the same locality, suggesting a mass mortality event and/or gregarious behaviour, and allowing the potential for studies of sexual dimorphism and population structure³¹. *Sapeornis* is the largest Early Cretaceous bird and its extremely elongated forelimbs suggest that it may have been a soaring form⁶⁴. Enantiornithine birds are common and are represented by the basal form *Protopteryx*³², with its unusual scale-like tail feathers, and *Longipteryx*, which possessed much longer wings than other known enantiornithines, which suggest a rather different ecology⁶⁵. Ornithurine birds are generally diverse but numerically rare, and included the basalmost ornithurine *Liaoningornis*, a small perching form, and the larger *Yixianornis* and *Yanornis*⁶⁶. The latter possessed an advanced flight apparatus, including an elongated, deeply keeled sternum and a coracoid of ‘modern’ appearance, both of which indicate strong flight capability (Figs 3 and 4). The presence of an intact seed mass in the preserved stomach contents of *Jeholornis* provides a specific instance of plant–animal interaction and indicates that frugivory appeared early in avian history⁶⁷.

The Jehol Group also yields important evidence on the evolutionary history of Cretaceous mammals. Six genera are known,

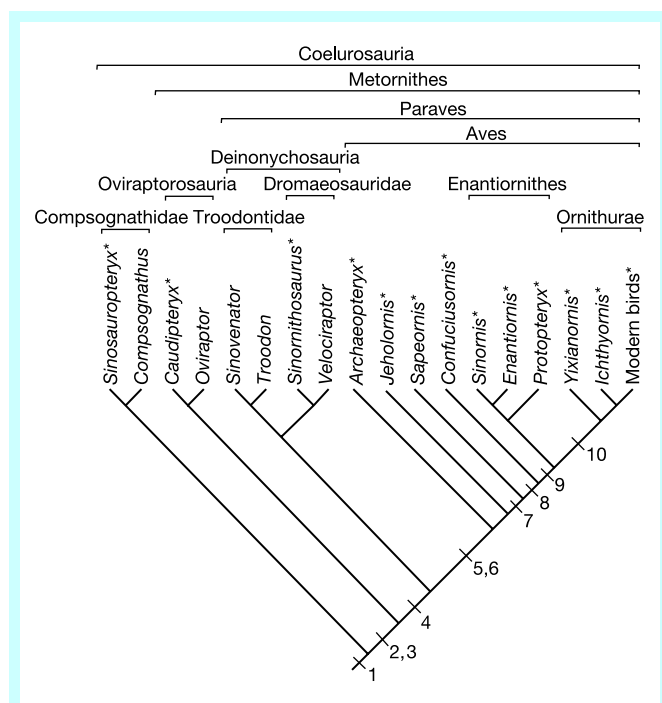


Figure 3 Cladogram illustrating the relationship of birds with major clades of non-avian coelurosaurian theropods (based on refs 34, 64, 67). The numbers in circles at each branching node indicate the first appearance of key morphological characters. 1, unbranched feathers; 2, uncinate processes on ribs; 3, true branched feathers; 4, retroverted pubis; 5, reversed hallux; 6, asymmetrical flight feathers; 7, pygostyle; 8, horny beak; 9, alula (bastard wing); 10, large, keeled sternum. Taxa indicated with an asterisk are known to have possessed either protofeathers or true feathers.

including basal forms (the eutriconodont *Jeholodens*⁶⁸ and the gobiconodontids *Repenomamus*^{38,39} and *Gobiconodon* sp.³⁸) and more derived taxa (the symmetrodont (basal therian) *Zhangheotherium*¹⁵, the multituberculate *Sinobaatar*⁶⁹ and the basal eutherian *Eomaia*³³). Each genus provides many character states amenable to phylogenetic analyses that in turn have elucidated sequences of character acquisition in Theria and Eutheria. For example, *Jeholodens* exhibits a mosaic of therian and non-therian character states⁶⁸, *Repenomamus* provides new information on the evolution of the mammalian middle ear and feeding apparatus³⁸, and *Zhangheotherium* demonstrates that a parasagittal gait was not characteristic of basal therians¹⁵. Of equal importance, the occurrence of *Eomaia* extends the range of eutherians into the Early Cretaceous, and indicates a pre-Barremian origin for the group that is much earlier than the origination dates posited by previous palaeontological and molecular clock studies³³. These fossils also offer a rare glimpse into the functional morphology and palaeoecology of Cretaceous mammals, which are usually known solely from dental material. *Eomaia* has limb and phalangeal proportions that are characteristic of extant arboreal and scansorial mammals³³, whereas *Zhangheotherium* and *Jeholodens* lack these specializations and were probably ground dwellers^{15,68}. The presence of epipubic bones in *Eomaia* suggests that it carried its young in a pouch, much like extant marsupials³³. The dentitions of all six taxa are consistent with insectivory or carnivory, and the animals ranged in size from *Jeholodens* (snout–vent length of about 50 mm)⁶⁸ to *Repenomamus* (skull length of 110 mm)³⁸, one of the largest Mesozoic mammals. These differences in body size and locomotory features indicate that mammals had diversified into a number of ecological roles by the late Early Cretaceous.

Invertebrate palaeontology

Invertebrates are the most abundant elements of the biota (Fig. 5)⁴. Nevertheless, although several groups have been studied intensively by biostratigraphers and taxonomists (for example, conchostracans), this work has had only limited effects on more general phylogenetic and palaeobiological issues. Much of the invertebrate material has yet to be described in detail and remains poorly known. Preliminary studies indicate that the diverse invertebrate fauna (Fig. 5) contained insects (Ephemeroptera, Odonata, Blattoidea, Hemiptera and Diptera)^{4,70,71}, spiders, crustaceans (Ostracoda, Notostraca, Conchostraca, Decapoda, Peracarida)^{4,72}, bivalves⁴ and gastropods⁴. The discovery of brachyceran flies with long tubular mouthparts suggests the presence of nectar-producing angiosperms in the Jehol flora⁷¹, provides circumstantial support for the suggestion that pollinating insects might have had an important role in the origin and early evolution of flowering plants, and may suggest long-lived plant–insect associations^{71,73}.

Palaeobotany

The Jehol flora is abundant, diverse and conifer dominated. It also contains bryophytes, lycopsids, sphenopsids, ferns, bennettitaleans, czekanowskialeans, ginkgoaleans, gnetaleans^{4,21} and enigmatic plants including putative early angiosperms^{16,21,74}. The latter, although rare, have been proposed as the oldest flowering plants^{16,21,74} and have received disproportionate attention owing to their possible relevance to the origin and early radiation of angiosperms. Most of these ‘angiosperms’ have now been discredited²¹, and only *Archaeofructus*^{16,21,74} is currently thought to represent a stem-group flowering plant. In our view, however, the affinities of this taxon remain controversial.

Archaeofructus contains three species: *A. liaoningensis*^{16,21,74}, *A. sinensis*⁷⁴ and a third unnamed species²¹. ‘Flowers’ of *Archaeofructus* contain male and female reproductive organs on the same shoot (Fig. 6). Male organs comprise paired, bilaterally symmetrical shoot stamens with slender filaments and long anthers, apparently containing two pollen sacs with monosulcate pollen. Stamens occur

below the female organs, mature first, and are abscised through ontogeny^{21,74}. Female organs are interpreted as stalked conduplicate carpels with adaxially elongated stigmatic crests. Carpels contain numerous obliquely oriented ovules (*A. liaoningensis*, 2–4; *A. sinensis*, 8–12)^{21,74}. Leaves are pinnately compound, highly divided and may be abscised^{21,74}. *Archaeofructus* lacked petals, sepals and other organs associated with stamens and carpels^{16,21,74}. Its flowers, therefore, are unlike those of both extant basal angiosperms^{74,75} and those Lower Cretaceous basal forms (in some cases referable to Nymphaeales) that possess small, compact flowers^{75,76}. The absence of these floral features, noted in most *bona fide* angiosperms^{75,76}, demonstrates that *Archaeofructus* does not belong within the angiosperm crown group, although it is conceivable that it lies on the angiosperm stem lineage. *Archaeofructus* shares several similarities with other Mesozoic seed plants such as *Caytonia*⁷⁴, from which it is apparent that its inferred phylogenetic position remains weakly supported⁷⁷. This problem is compounded by a lack of unambiguous anatomical data supporting key features in *Archaeofructus*. Specifically, we regard the presence of two pollen sacs in each of the thecae as unproven, and although it was suggested that the ovules of *Archaeofructus* were enclosed in a carpel^{16,21,74}, the description of this structure lacks sufficient detail to identify synapomorphic characters that would link it specifically with angiosperms. This is important as carpel-like structures are known in other seed plants that are distantly related to angiosperms, such as caytoniales^{75,78,79}. On the basis of available published data, it is currently impossible to determine whether the ovule-enclosing structure in *Archaeofructus* is a carpel, or merely carpel-like. Several other important features that would influence the phylogenetic position of *Archaeofructus* (such as cuticular structure, shoot anatomy

and ovule histology) are also unknown. Therefore, we cannot unquestioningly accept *Archaeofructus* as an angiosperm.

If *Archaeofructus* is a stem-group angiosperm, the Late Hauterivian/Early Aptian age of the deposits indicates that it significantly post-dates the first appearance of crown-group angiosperms⁵. Consequently, *Archaeofructus* cannot represent the oldest known flowering plant. It would, however, offer insights into the origin of the crown group and of the character states associated with it. For example, its morphology suggests it to have been herbaceous and aquatic, indicating that flowering plants might have evolved in riparian environments where competition with other seed plants was, perhaps, less intense^{74,77}. The apparently paired stamens in *Archaeofructus* could indicate that basal flowering plants bore male and female organs on separate shoots and that these later evolved into shorter bisexual flowers^{74,77}, and that stamens are remnants of earlier branching systems⁷⁴. Ideally, *Archaeofructus* needs to be included in more exhaustive cladistic analyses with wider taxon sampling of fossil plants (Bennettitales, Caytoniales and extinct Gnetales) that are known to be of relevance to angiosperm origin^{75,79}, and using a greater range of morphological characters⁷⁹. Although *Archaeofructus* may be the sister clade of crown-group angiosperms⁷⁴, the possibility that it represents a previously unrecognized group of seed plants cannot at this stage be excluded.

Non-angiosperms in the flora have received little attention so far, but several studies have attempted to place the Jehol plants within wider biostratigraphical, regional and phylogenetic contexts. For example, Jehol gnetaleans contribute significantly to our knowledge of the evolutionary history of this paucispecific group. These plants possess features that allow them to be affiliated with extant genera: for example, three *Ephedrites* species (see Fig. 6 for an example)



Figure 4 Vertebrate fossils from the Jehol Biota, illustrating key morphological changes during the transition from non-avian coelurosaurs to modern birds. Numbering of the following figures corresponds to the sequence of acquisition of the various character states shown on the cladogram in Fig. 3: 1, unbranched, filamentous integumentary structures associated with the tail of *Sinosauropteryx* sp. (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing (IVPP) V12415); 2, *Caudipteryx dongi* (IVPP V12344), showing the presence of uncinat

dorsal ribs; 3, feathers of modern, branched appearance from *Caudipteryx* sp. (IVPP V12430); 4, retroverted pubis of *Sinovenator zhangii* (IVPP V12615); 5, pes of *Confuciusornis sanctus* (IVPP V12352) showing reversed hallux; 6, asymmetrical flight feather of *Archaeopteryx lithographica* (Tithonian, Germany); 7, pygostyle of *Sapeornis chaoyangensis* (IVPP V13275); 8, *Confuciusornis sanctus*, showing preservation of a horny keratinous beak (IVPP V12352); 9, the alula of *Protopteryx fengningensis* (IVPP V11665); 10, keeled sternum of *Yixianornis grabi* (IVPP V12631).

are closely related to living *Ephedra*²¹. Similarly, the Jehol form *Gurvanella* (see Fig. 6 for an example) is reproductively similar to extant *Welwitschia*²¹, but resembles extant *Ephedra* in its vegetative features. These observations support a closer relationship between *Ephedra* and *Welwitschia* than usually proposed⁸⁰, and suggest a relatively recent, pre-Barremian, divergence of families including extant genera⁷⁹. The vegetative morphology of *Gurvanella* (Fig. 6) is typical of most gnetales, but differs radically from that of closely related extant *Welwitschia*, supporting the hypothesis that the latter is highly derived and adapted to extreme environmental conditions. Most Jehol conifers represent typical Mesozoic forms (Fig. 6), although some have been considered to be equivalents of living genera (*Podocarpites*, *Araucarites*^{4,21}). Similarly, the Jehol sphenop-sid *Equisetites* bears a strong resemblance to the living horsetail *Equisetum*^{4,21}. In both cases, the similarity of these plants to extant genera, in combination with the age of the deposits, suggests that Jehol plants bridge chronological and evolutionary gaps between Mesozoic plants and their living relatives. As a result, they should be able to contribute significantly to debates on conifer phylogeny and the origin of modern horsetails.

Palaeobiogeography: a refugium, a cradle or both?

Although radiometric dating and biostratigraphical correlations provide overwhelming evidence for an Early Cretaceous age, the

closest relatives of several Jehol taxa are from Late Jurassic, or older, deposits. This pattern has prompted the hypothesis that east Asia acted as a refugium for some of these more typically 'Jurassic' taxa in the Lower Cretaceous⁸¹. Such 'relicts' include the theropod dinosaur *Sinosauropteryx* (sister taxon of the Upper Jurassic *Compsognathus*) and the anurognathid pterosaur *Dendrorhynchoides* (no other post-Jurassic anurognathids are known currently). Palaeogeography and the composition of terrestrial vertebrate faunas provide some support for this hypothesis, as eastern Asia seems to have been isolated from the rest of Laurasia during Middle Jurassic to Early Cretaceous times^{82–84}. Potential isolating mechanisms included the Mongol–Okhotsk Sea between eastern Asia and Siberia, the epicontinental Turgai Sea between Europe and Central Asia, and breaches in the tenuous Junggar–Tarim–Mongolian land bridge, which had linked east Asia with the rest of Laurasia during the Triassic and Early Jurassic^{82,83}. The occurrence of a relict tritylodontid synapsid (a group thought to have become extinct at the end of the Middle Jurassic) in the Early Cretaceous of Japan⁸⁵ adds some support to the refugial hypothesis, as does the presence of several endemic plant and vertebrate clades in eastern Asia at this time, including *Archaeofructus*, psittacosaurid dinosaurs, testudinoid turtles, and sinamiid and peipiaosteoid freshwater fishes^{4,16,30,58,86,87}.

The palaeobiogeographical history of this region was complex, and the composition of the Jehol Biota is only partly explained by

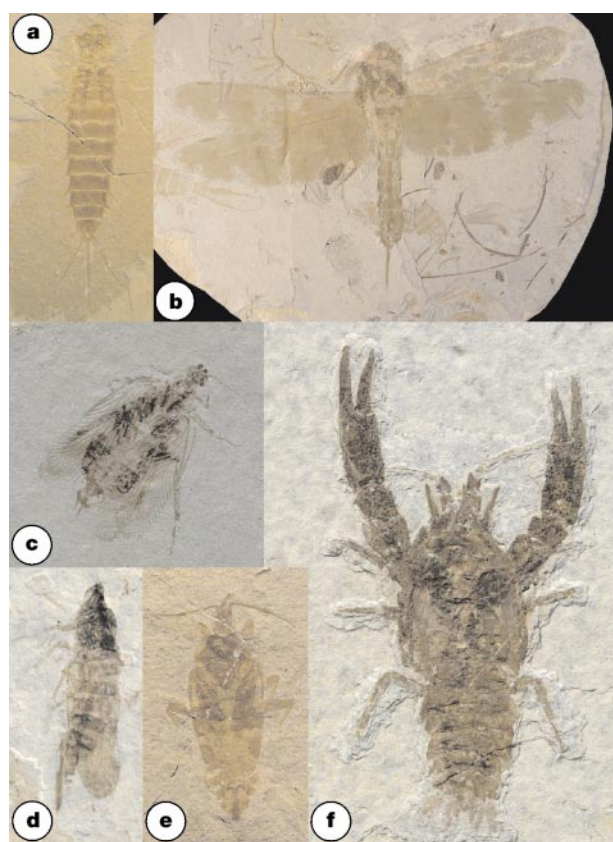


Figure 5 Selected invertebrates from the Jehol fauna. **a**, A larval specimen of the biozonal ephemeropter insect *Ephemeropsis trisetalis* (IVPP I0237). **b**, Adult specimen of *Aeschnidium heishankowense* (Insecta: Odonata) (Nanjing Institute of Geology and Palaeontology (NIGP) B930042-2). **c**, Undescribed cockroach (Blattoidea) referable to the Mesoblattinidae (IVPP I0076). **d**, An undescribed proceropod (Hemiptera) insect (IVPP I0050). **e**, Hemipteran insect (*Mesanthocoris* sp.: NIGP, uncatalogued specimen), note the well-preserved stylus. **f**, The crayfish *Cricoidoselosus aethus* (IVPP I0041).



Figure 6 Selected plants from the Jehol flora. **a**, Fertile shoot of the gnetalean *Ephedrites* sp. (IVPP B0129). **b**, Fertile branching system of the gnetalean *Gurvanella* (holotype of *Chaoyangia liangii* Duan 1998; specimen 9341). **c**, Branching structure of *Gurvanella* sp. demonstrating ribbed stem, nodal architecture and small leaves originating at the level of branching (arrowed), consistent with gnetalean affinities. **d**, Seed of *Gurvanella* sp. with delicate wing and *Welwitschia*-like morphology. **e**, Fertile shoot of the putative basal angiosperm *Archaeofructus liaoningensis* with ovulate units distal on the shoot, and male organs proximal (IVPP B0161). **f**, Broad-leaved conifer, *Cephalotaxopsis* sp. (IVPP B0048).

the refugium hypothesis⁸⁸. In addition to the various relict and endemic taxa listed above, east Asian Lower Cretaceous faunas also contained early, and in many cases basal, representatives of dinosaur clades that are more often associated with Upper Cretaceous biomes (for example, tyrannosauroid⁸⁹, therizinosauroid and oviraptorosaurian theropods^{14,25,37,88}, and neoceratopians³⁵), suggesting that this region could also be regarded as a centre of diversification for some of these taxa⁸⁸. These faunas also include members of several clades that have a more cosmopolitan Early Cretaceous distribution, including various dinosaurs (iguanodontian ornithopods⁵⁷, titanosauriform sauropods⁹⁰, nodosaurid ankylosaurs⁵⁹, dromaeosaurid theropods^{26–29}), discoglossid frogs^{50,91}, paramacellodid lizards⁹¹, ctenochasmatid pterosaurs⁵⁵, enantiornithine birds^{32,65} and an eobaatarid multituberculate mammal⁶⁹. The flora is closely comparable with that from other Lower Cretaceous sites in Mongolia and northern China^{92,93} and is broadly similar to contemporaneous floras from other regions. In particular, the Jehol gnetaleans are very similar to those from coeval formations in southern Russia²¹ and Brazil⁹⁴. Indeed, in these respects the Jehol Biota can be regarded as a 'typical' Early Cretaceous biome, and is broadly comparable to penecontemporaneous terrestrial biomes from western Europe, North and South America, Africa and Australia. Most palaeobiogeographical models suggest that Asian isolation ended close to the end of the Early Cretaceous, sometime during either the Aptian or Albian stages^{82–84}. However, the presence of titanosauriform, dromaeosaurid and iguanodontian dinosaurs in the Jehol Biota and in other east Asian faunas (from the Berriasian/Barremian of China, Japan, Korea and Thailand) indicates that isolation must have ended earlier in the Cretaceous, perhaps in the Berriasian⁹⁰.

Why did east Asia host a combination of relicts, cosmopolitan taxa and endemics at this time? The temporal proximity of the Jehol Biota to the time when east Asian isolation ended may offer a partial explanation for the mixed composition of the biome. The presence of relict taxa may simply reflect the long period of isolation through the Middle Jurassic to the earliest Cretaceous, whereas the arrival of more cosmopolitan forms probably signalled the breaching of various palaeogeographical barriers (such as regression of the Turgai Sea)^{82–84,88,90}. The Jehol Group can be viewed as a window on succession in an Early Cretaceous terrestrial biome, in which an established biota merged with and was partially replaced by a novel biota composed of immigrants and new taxa that were evolving *in situ*. The high degree of endemism is probably explicable in terms of east Asian isolation generally and, more specifically, in the isolation of the Jehol Biota within eastern Asia by additional local palaeogeographical barriers (such as the Qing and Dabieshan Mountains⁷).

Collection strategies and associated problems

Advances in our understanding of the Jehol Biota have been hampered by illegal collecting⁹⁵, manufacture of faked and composite specimens (as in the 'Archaeoraptor' debacle^{95,96}), illegal sale and export of fossils^{97,98}, and difficulties in the acquisition of specimens by international scientific institutions bound by ethical collecting standards⁹⁸. Currently, the main problem for legitimate scientific investigation is collection by private individuals. Such collectors do not record the contextual information of locality, stratigraphy and sedimentology essential for placing the organisms within their palaeoenvironmental and taphonomic settings. To bypass these problems, and to advance research on this important biome, excavations sponsored by research institutes, under the supervision and guidance of local authorities, must become the norm. Systematic, responsible collecting and vigilant site management are needed to stop the loss of important and unique specimens.

Conclusions

The spectacular fossils of the Jehol Group have already provided many important insights into the evolution of birds, angiosperms

and mammals. Nevertheless, the rate of fossil discovery presently outstrips the rate of description, and detailed monographic treatments of all species from the biota are needed if the full potential of these deposits is to be realized. The Jehol Biota currently represents our best chance of viewing the composition and dynamics of an intact Early Cretaceous terrestrial ecosystem: continuing study of the fauna, flora, taphonomy and palaeoenvironment is likely to yield exciting new results for years to come.

Note added in proof: New material of the dromaeosaurid theropod *Microraptor* indicates that this animal possessed wing-like arrays of asymmetrical feathers¹⁰⁰ on both its fore- and hindlimbs. This observation provides additional support for the hypothesis that this taxon was an arboreal glider. □

doi:10.1038/nature01420.

- Grabau, A. W. *Stratigraphy of China Part 2 Mesozoic* (Geological Survey of China, Beijing, 1928).
- Chen, P.-J. *et al.* The late Mesozoic continental formations of western Liaoning. *Bull. Nanjing Inst. Geol. Palaeontol.* **1**, 22–55 (1980). (In Chinese with English summary.)
- Chen, P.-J. Distribution and migration of the Jehol Fauna with reference to non-marine Jurassic–Cretaceous boundary in China. *Acta Palaeontol. Sin.* **27**, 659–683 (1988). (In Chinese with English summary.)
- Chen, P.-J. & Jin, F. (eds). The Jehol Biota. *Palaeoworld* **11**, 1–342 (1999). (In Chinese with English summary.)
- Barrett, P. M. Evolutionary consequences of dating the Yixian Formation. *Trends Ecol. Evol.* **15**, 99–103 (2000).
- Wang, X.-L. *et al.* Vertebrate faunas and biostratigraphy of the Jehol Group in western Liaoning, China. *Vert. Palaeontol.* **38** (Suppl.), 41–63 (2000).
- Zhang, M., Chen, P.-J., Wang, Y.-Q. & Wang, Y. *The Jehol Biota* (Shanghai Science and Technology, Shanghai, 2001). (In Chinese.)
- Smith, J. B., Harris, J. D., Omar, G. I., Dodson, P. & You, H.-L. in *New Perspectives on the Origin and Early Evolution of Birds* (eds Gauthier, J. & Gall, L. F.) 549–589 (Peabody Museum of Natural History, New Haven, 2001).
- Hou, L.-H., Zhou, Z.-H., Gu, Y.-C. & Zhang, H. *Confuciusornis sanctus*, a new Late Jurassic sauriurine bird from China. *Chin. Sci. Bull.* **40**, 1545–1551 (1995). (In Chinese.)
- Hou, L.-H., Zhou, Z.-H., Martin, L. D. & Feduccia, A. A beaked bird from the Jurassic of China. *Nature* **377**, 616–618 (1995).
- Ji, Q. & Ji, S.-A. On discovery of the earliest bird fossil in China and the origin of birds. *Chin. Geol.* **10**, 30–33 (1996). (In Chinese.)
- Ji, Q. & Ji, S.-A. *Protarchaeopteryx*, a new genus of Archaeopterygidae in China. *Chin. Geol.* **3**, 38–41 (1997). (In Chinese.)
- Chen, P.-J., Dong, Z.-M. & Zhen, S.-N. An exceptionally well-preserved theropod dinosaur from the Yixian Formation of China. *Nature* **391**, 147–152 (1988).
- Ji, Q., Currie, P. J., Norell, M. A. & Ji, S.-A. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
- Hu, Y.-M., Wang, Y.-Q., Luo, Z.-X. & Li, C.-K. A new symmetrodont mammal from China and its implications for mammalian evolution. *Nature* **390**, 137–142 (1997).
- Sun, G., Dilcher, D. L., Zheng, S.-L. & Zhou, Z.-K. In search of the first flower: a Jurassic angiosperm, *Archaeofructus*, from north-east China. *Science* **282**, 1692–1695 (1998).
- Wang, W.-L. *et al.* Mesozoic stratigraphy and palaeontology of western Liaoning, China (Chinese Geological Press, Beijing, 1989). (In Chinese with English summary.)
- Smith, A. G., Smith, D. G. & Funnell, B. M. *Atlas of Mesozoic and Cenozoic Coastlines* (Cambridge Univ. Press, Cambridge, 1994).
- Allison, P. A. & Briggs, D. E. G. (eds) *Taphonomy: Releasing the Data Locked in the Fossil Record* (Plenum, New York/London, 1991).
- Wang, H.-Z., Yang, S.-N. & Li, S.-T. Mesozoic and Cenozoic basin formation in east China and adjacent regions and development of the continental margin. *Acta. Geol. Sinica* **57**, 213–223 (1983).
- Sun, G., Zheng, S.-L., Dilcher, D. L., Wang, Y.-D. & Mei, S.-W. *The Early Angiosperms and their Associated Plants from Western Liaoning, China* (Shanghai Scientific and Technological Education, Shanghai, 2001). (In Chinese with English summary.)
- Wang, X.-L. *et al.* Stratigraphic sequence and vertebrate-bearing beds of the lower part of the Yixian Formation in Sihetun and neighbouring area, western Liaoning, China. *Vert. Palaeontol.* **36**, 81–101 (1998).
- Wang, X.-L. *et al.* Record of the Sihetun vertebrate mass mortality events, western Liaoning, China: caused by volcanic eruptions. *Geol. Rev.* **45** (Suppl.), 458–466 (1999). (In Chinese.)
- Wang, X.-L., Zhou, Z.-H., Zhang, F.-C. & Xu, X. A nearly completely articulated rhamphorhynchoid pterosaur with exceptionally well-preserved wing membranes and 'hairs' from Inner Mongolia, northeast China. *Chin. Sci. Bull.* **47**, 226–230 (2002).
- Xu, X., Tang, Z.-L. & Wang, X.-L. A therizinosauroid dinosaur with integumentary structures from China. *Nature* **399**, 350–354 (1999).
- Xu, X., Wang, X.-L. & Wu, X.-C. A dromaeosaurid dinosaur with a filamentous integument from the Yixian Formation of China. *Nature* **401**, 262–266 (1999).
- Xu, X., Zhou, Z.-H. & Prum, R. O. Branched integumentary structures in *Sinornithosaurus* and the origin of feathers. *Nature* **410**, 200–204 (2001).
- Ji, Q., Norell, M. A., Gao, K.-Q., Ji, S.-A. & Ren, D. The distribution of integumentary structures in a feathered dinosaur. *Nature* **410**, 1084–1088 (2001).
- Norell, M. A. *et al.* 'Modern' feathers on a non-avian dinosaur. *Nature* **416**, 36–37 (2002).
- Mayr, G., Peters, D. S., Plodowski, G. & Vogel, O. Bristle-like integumentary structures at the tail of the horned dinosaur *Psittacosaurus*. *Naturwissenschaften* **89**, 361–365 (2002).
- Chiappe, L. M., Ji, S.-A., Ji, Q. & Norell, M. A. Anatomy and systematics of the Confuciusornithidae (Theropoda: Aves) from the Late Mesozoic of northeastern China. *Bull. Am. Mus. Nat. Hist.* **242**, 1–89 (1999).

32. Zhang, F.-C. & Zhou, Z.-H. A primitive enantiornithine bird and the origin of feathers. *Science* **290**, 1955–1959 (2000).
33. Ji, Q. *et al.* The earliest known eutherian mammal. *Nature* **416**, 816–822 (2002).
34. Xu, X., Norell, M. A., Wang, X.-L., Makovicky, P. J. & Wu, X.-C. A basal troodontid from the Early Cretaceous of China. *Nature* **415**, 780–784 (2002).
35. Xu, X., Makovicky, P. J., Wang, X.-L., Norell, M. A. & You, H.-L. A ceratopsian dinosaur from China and the early evolution of Ceratopsia. *Nature* **416**, 314–317 (2002).
36. Xu, X., Wang, X.-L. & You, H.-L. A primitive ornithomimid from the Early Cretaceous Yixian Formation of Liaoning. *Vert. Palasiat.* **38**, 318–325 (2000).
37. Xu, X., Cheng, Y.-N., Wang, X.-L. & Chang, C.-H. An unusual oviraptorosaurian dinosaur from China. *Nature* **419**, 291–293 (2002).
38. Wang, Y.-Q., Hu, Y.-M., Meng, J. & Li, C.-K. An ossified Meckel's cartilage in two Cretaceous mammals and origin of the mammalian middle ear. *Science* **294**, 357–361 (2001).
39. Li, J.-L., Wang, Y., Wang, Y.-Q. & Li, C.-K. A new family of primitive mammal from the Mesozoic of western Liaoning, China. *Chin. Sci. Bull.* **46**, 782–785 (2001).
40. Jiang, F.-X., Chen, J.-H., Cao, M.-Z. & Komatsu, T. A discussion on the age of the feathered dinosaurs-bearing beds of Liaoning, China. *Acta Palaeontol. Sinica* **39**, 307–311 (2000). (In Chinese with English summary.)
41. Swisher, C. C. III *et al.* Cretaceous Age for the feathered dinosaurs of Liaoning, China. *Nature* **400**, 58–61 (1999).
42. Swisher, C. C. III *et al.* Further Support for a Cretaceous age for the feathered-dinosaur beds of Liaoning, China: New ⁴⁰Ar/³⁹Ar Dating of the Yixian and Tuchengzi Formations. *Chin. Sci. Bull.* **47**, 135–138 (2002).
43. Li, P.-X., Su, D.-Y., Li, Y.-G. & Yu, J.-X. Age assignment of the *Lycoptera*-bearing beds. *Acta Geol. Sinica* **7**, 329–349 (1994). (English edn.)
44. Wang, D.-F. & Diao, N.-C. *Scientific Papers on Geology for International Exchange* Vol. 1, 1–12 (Geological Publishing House, Beijing, 1984).
45. Lo, C.-H., Chen, P.-J., Tsou, T.-Y., Sun, S.-S. & Lee, C.-Y. ⁴⁰Ar–³⁹Ar laser single-grain and K–Ar dating of the Yixian Formation, NE China. *Palaeoworld* **11**, 329–340 (1999).
46. Wang, S.-S., Hu, H.-G., Li, P.-X. & Wang, Y.-Q. Further discussion on geologic age of Sihetun vertebrate assemblage in western Liaoning, China: evidence from Ar–Ar dating. *Act. Petrolog. Sinica* **17**, 663–668 (2001). (In Chinese.)
47. Smith, P. E. *et al.* Dates and rates in ancient lakes: ⁴⁰Ar–³⁹Ar evidence for an Early Cretaceous age for the Jehol Group, Northeast China. *Can. J. Earth Sci.* **32**, 1426–1431 (1995).
48. Eberth, D. A., Russell, D. A., Braman, D. R. & Deino, A. L. The age of the dinosaur-bearing sediments at Tebch, Inner Mongolia, People's Republic of China. *Can. J. Earth Sci.* **30**, 2101–2106 (1993).
49. Grande, L., Jin, F., Yabumoto, Y. & Bemis, W. E. *Protopsephurus liui*, a well-preserved primitive paddlefish (Acipenseriformes: Polyodontidae) from the Lower Cretaceous of China. *J. Vert. Paleontol.* **22**, 209–237 (2002).
50. Wang, Y., Gao, K.-Q. & Xu, X. Early evolution of discoglossid frogs: new evidence from the Mesozoic of China. *Naturwissenschaften* **87**, 417–420 (2000).
51. Gao, K.-Q. & Shubin, N. H. Late Jurassic salamanders from northern China. *Nature* **410**, 574–577 (2001).
52. Gao, K.-Q., Evans, S. E., Ji, Q., Norell, M. A. & Ji, S.-A. Exceptional fossil material of a semi-aquatic reptile from China: the resolution of an enigma. *J. Vert. Paleontol.* **20**, 417–421 (2000).
53. Smith, J. B. & Harris, J. D. A taxonomic problem concerning two diapsid genera from the Lower Yixian Formation of Liaoning Province, northeastern China. *J. Vert. Paleontol.* **21**, 389–391 (2001).
54. Endo, R. & Shikama, T. Mesozoic reptilian fauna in the Jehol mountainland, Manchoukuo. *Bull. Cent. Nat. Mus. Manchoukuo* **3**, 1–20 (1942).
55. Unwin, D. M., Lu, J.-C. & Bakurina, N. N. On the systematic and stratigraphical significance of pterosaurs from the Lower Cretaceous Yixian Formation (Jehol Group) of Liaoning, China. *Mitt. Mus. Nat. kd. Berl. Geowiss. Reihe* **3**, 181–206 (2000).
56. Wang, X.-L. & Lu, J.-C. Discovery of a pterodactylid pterosaur from the Yixian Formation of western Liaoning, China. *Chin. Sci. Bull.* **46**, 1112–1117 (2001).
57. Wang, X.-L. & Xu, X. A new iguanodontid (*Jinzhousaurus yangi* gen. et sp. nov.) from the Yixian Formation of western Liaoning, China. *Chin. Sci. Bull.* **46**, 1669–1672 (2001).
58. Xu, X. & Wang, X.-L. New psittacosaur (Ornithischia, Ceratopsia) occurrence from the Yixian Formation of Liaoning, China and its stratigraphical significance. *Vert. Palasiat.* **36**, 147–158 (1998).
59. Xu, X., Wang, X.-L. & You, H.-L. A juvenile ankylosaur from China. *Naturwissenschaften* **88**, 297–300 (2001).
60. Padian, K. Cross-testing adaptive hypotheses: phylogenetic analysis and the origin of bird flight. *Am. Zool.* **41**, 598–607 (2001).
61. Padian, K., Ji, Q. & Ji, S.-A. in *Mesozoic Vertebrate Life* (eds Tanke, D. H., Carpenter, K. & Skrepnick, M. W.) 117–135 (Indiana Univ. Press, Bloomington, 2001).
62. Martin, L. D. & Czerkas, S. A. The fossil record of feather evolution in the Mesozoic. *Am. Zool.* **40**, 687–694 (2000).
63. Xu, X., Zhou, Z.-H. & Wang, X.-L. The smallest known non-avian theropod dinosaur. *Nature* **408**, 705–708 (2000).
64. Zhou, Z.-H. & Zhang, F.-C. Largest bird from the Early Cretaceous and its implications for the earliest avian ecological diversification. *Naturwissenschaften* **89**, 34–38 (2002).
65. Zhang, F.-C., Zhou, Z.-H., Hou, L.-H. & Gu, G. Early diversification of birds: evidence from a new opposite bird. *Chin. Sci. Bull.* **46**, 945–949 (2001).
66. Zhou, Z.-H. & Zhang, F.-C. Two new ornithurine birds from the Early Cretaceous of western Liaoning, China. *Chin. Sci. Bull.* **46**, 1258–1264 (2001).
67. Zhou, Z.-H. & Zhang, F.-C. A long-tailed, seed-eating bird from the Early Cretaceous of China. *Nature* **418**, 405–409 (2002).
68. Ji, Q., Luo, Z.-X. & Ji, S.-A. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398**, 326–330 (1999).
69. Hu, Y.-M. & Wang, Y.-Q. *Sinobaatar* gen. nov.: first multituberculata from the Jehol Biota of Liaoning, northeast China. *Chin. Sci. Bull.* **47**, 933–938 (2002).
70. Ren, D. in *Fauna and Stratigraphy of Jurassic-Cretaceous in Beijing and the Adjacent Areas* (eds Ren, D., Lu, L.-W., Guo, Z.-G. & Ji, S.-A.) 47–121 (Seismology, Beijing, 1995). (In Chinese with English summary.)
71. Ren, D. Flower-associated Brachycera flies as fossil evidence for Jurassic angiosperm origins. *Science* **280**, 85–88 (1998).
72. Taylor, R. S., Schram, F. R. & Shen, Y.-B. A new crayfish Family (Decapoda: Astacidae) from the Upper Jurassic of China, with a reinterpretation of other Chinese crayfish taxa. *Paleontol. Res.* **3**, 121–136 (1999).
73. Labandeira, C. C., Dilcher, D. L., Davis, D. R. & Wagner, D. L. Ninety-seven million years of angiosperm-insect association: paleobiological insights into the meaning of coevolution. *Proc. Natl. Acad. Sci. USA* **91**, 12278–12282 (1994).
74. Sun, G. *et al.* Archaeofractaceae, a new basal angiosperm family. *Science* **296**, 899–904 (2002).
75. Crane, P. R., Friis, E. M. & Pedersen, K. R. The origin and early diversification of angiosperms. *Nature* **374**, 27–33 (1995).
76. Friis, E. M., Pedersen, K. R. & Crane, P. R. Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* **401**, 357–360 (2001).
77. Stokstad, E. Fossil plant hints how first flower bloomed. *Science* **296**, 821 (2002).
78. Harris, T. M. *The Yorkshire Jurassic Flora. II. Caytoniales, Cycadales and Pteridosperms* (British Museum of Natural History, Oxford, 1964).
79. Rothwell, G. W. & Serbet, R. Lignophyte phylogeny and the evolution of spermatophytes: a numerical cladistic analysis. *Syst. Bot.* **19**, 443–482 (1994).
80. Rydin, C., Källersjö, M. & Friis, E. M. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems, and the monophyly of conifers. *Int. J. Plant Sci.* **163**, 197–214 (2002).
81. Luo, Z.-X. A refugium for relicts. *Nature* **400**, 24–25 (1999).
82. Russell, D. A. The role of Central Asia in dinosaurian biogeography. *Can. J. Earth Sci.* **30**, 2002–2012 (1993).
83. Upchurch, P. The evolutionary history of the sauropod dinosaurs. *Phil. Trans. R. Soc. Lond. B* **349**, 365–390 (1995).
84. Upchurch, P., Huxley, C. A. & Norman, D. B. An analysis of dinosaurian biogeography: evidence for the existence of vicariance and dispersal patterns caused by geological events. *Proc. R. Soc. Lond. B* **269**, 613–621 (2002).
85. Setoguchi, T., Matsuoka, H. & Matsuda, M. in *Proc. 7th Annu. Meeting Chin. Soc. Vert. Paleontol.* (eds Wang, Y.-Q. & Deng, T.) 117–124 (China Ocean, Beijing, 1999).
86. Hirayama, R., Brinkman, D. B. & Danilov, I. G. Distribution and biogeography of non-marine Cretaceous turtles. *Russian J. Herpetol.* **7**, 181–198 (2000).
87. Grande, L. & Bemis, W. E. A comprehensive phylogenetic study of amii fishes (Amiidae) based on comparative skeletal anatomy. An empirical search for interconnected patterns of natural history. *Mem. Soc. Vert. Paleontol.* **4**, 1–690 (1998).
88. Manabe, M., Barrett, P. M. & Isaji, S. A refugium for relicts? *Nature* **404**, 953 (2000).
89. Manabe, M. The early evolution of the Tyrannosauridae in Asia. *J. Paleontol.* **73**, 1176–1178 (1999).
90. Barrett, P. M., Hasegawa, Y., Manabe, M., Isaji, S. & Matsuoka, H. Sauropod dinosaurs from the Lower Cretaceous of Eastern Asia: taxonomic and biogeographical implications. *Palaeontology* **45**, 1197–1217 (2002).
91. Evans, S. E. *et al.* An Early Cretaceous assemblage from Gifu Prefecture, Japan. *N. Mex. Mus. Nat. Hist. Sci. Bull.* **14**, 183–186 (1998).
92. Deng, S.-H. *Early Cretaceous Flora of Hulinhe Basin, Inner Mongolia, Northeast China* (Geological Publishing House, Beijing, 1995). (In Chinese with English summary.)
93. Deng, S.-H., Ren, S. & Fen, C. *Early Cretaceous Floras of Hailar, Inner Mongolia, China* (Geological Publishing House, Beijing, 1997). (In Chinese with English summary.)
94. Martill, D. M. *Fossils of the Santana and Crato Formations, Brazil* (Palaeontological Association, London, 1993).
95. Dalton, R. Feathers fly over Chinese fossil birds legality and authenticity. *Nature* **403**, 689–690 (2000).
96. Rowe, T. *et al.* The *Archaeoraptor* forgery. *Nature* **410**, 539–540 (2001).
97. Dalton, R. Chasing the dragons. *Nature* **406**, 930–932 (2000).
98. Dalton, R. Wandering Chinese fossil turns up at museum. *Nature* **414**, 571 (2001).
99. Enkin, R. J., Yang, Z.-Y., Chen, Y. & Courtillot, V. Paleomagnetic constraints on the geodynamic history of the major blocks of China from the Permian to the Present. *J. Geophys. Res.* **97**, 13953–13989 (1992).
100. Xu, X. *et al.* Four-winged dinosaurs from China. *Nature* **421**, 335–340 (2003).

Acknowledgements We thank the IVPP Liaoxi Project field team (M.-M. Zhang, X.-L. Wang, X. Xu, F. Zhang, Y. Wang, F. Jin, J.-Y. Zhang, Y.-Q. Wang, Y.-M. Hu), and C. C. Swisher, Q. Leng, H. Zhang, J. Zhang, C.-S. Li, S. Duan, W.-Q. Zhu, E.-M. Friis, J. Wocicki, P. G. Davis, P. Upchurch, D. M. Unwin, M. A. Taylor, R. Paton and I. Glasspool for discussions. K. Padian and Z.-X. Luo provided useful comments. This work was supported by the National Natural Science Foundation of China (to Z.Z.), the Chinese Academy of Sciences (to Z.Z.), the Special Funds for Major State Basic Research Projects of China (to Z.Z.), The Royal Society of London (to P.M.B.), Jurassic Foundation (to P.M.B.), and NERC (Research Fellowship to J.H.).

Correspondence and requests for materials should be addressed to Z.Z. (e-mail: zhonghe@yeah.net).

The importance of water to oceanic mantle melting regimes

P. D. Asimow* & C. H. Langmuir†

* Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

† Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

The formation of basaltic crust at mid-ocean ridges and ocean islands provides a window into the compositional and thermal state of the Earth's upper mantle. But the interpretation of geochemical and crustal-thickness data in terms of magma source parameters depends on our understanding of the melting, melt-extraction and differentiation processes that intervene between the magma source and the crust. Much of the quantitative theory developed to model these processes has neglected the role of water in the mantle and in magma, despite the observed presence of water in ocean-floor basalts. Here we extend two quantitative models of ridge melting, mixing and fractionation to show that the addition of water can cause an increase in total melt production and crustal thickness while causing a decrease in mean extent of melting. This may help to resolve several enigmatic observations in the major- and trace-element chemistry of both normal and hotspot-affected ridge basalts.

Much of the correlated variability among lava chemistry, crustal thickness, and axial depth observed along the global mid-ocean ridge system can be explained with reference to polybaric, anhydrous melting of peridotite^{1–4}. The principal variable in this melting process is mantle potential temperature^{3,4}, modified by effects such as conduction to the surface^{5–8}, geometry of mantle flow (passive versus buoyancy-driven circulation)^{7,9}, extent of equilibration between migrating melts and mantle^{2,7,10}, and possibly major-element source composition. However, certain geochemically and bathymetrically anomalous regions such as the Galápagos⁷ and Azores¹¹ platforms do not fit the global trends of variability, and are characterized by high concentrations of dissolved H₂O in erupted basalt glass¹². Here we demonstrate with two models how the significant effects of H₂O on melting¹³ and differentiation¹⁴ affect the mean properties that are calculated for the ocean-ridge melting regime. In contrast to anhydrous melting processes, H₂O in the source increases the total melt production (and hence crustal thickness) and mean pressure of melting while decreasing the mean extent of melting. This model accounts for many features of the anomalous regions, and also is able to resolve a conflict between variations in major and trace elements on normal ocean ridges.

Two useful properties that summarize mantle melt production beneath ocean ridges are the mean (or bulk) extent of melting, F_B (ref. 15), and the crustal thickness, Z_c . F_B is defined as the mass of melt generated per unit time over the whole melting regime divided by the mass flux of subsolidus source into the melting regime. It follows that the enrichment of a perfectly incompatible tracer in the aggregate primary melt over its source is $1/F_B$. Furthermore, for most steady-state melting regimes there is a residual mantle column¹⁶ (RMC) such that the crustal thickness in pressure units (which can be converted to kilometres from an estimate of crustal density) is given by F_B times the difference in pressure between the final pressure of melting (P_f) and the maximum pressure of residual material from which melt is extracted—which in many cases is the initial pressure of melting (P_o) (ref. 16). The ‘shape’ of the melting regime and the presence of active flow do not change the relation between F_B and Z_c (ref. 16), and only in cases where the depth to the solidus varies across axis, or melt is retained in some deep off-axis interval, is the connection to P_o obscured. For a model with near-constant productivity (that is, percentage melting per unit pressure drop), F_B increases linearly and Z_c quadratically with P_o , which in turn increases monotonically with potential temperature (T_p). Therefore F_B , mean pressure of melting ($\bar{P}$) and Z_c all correlate

positively³. Furthermore, most complications to this scheme, such as modifications to the productivity function, the extent of buoyancy-driven active flow, or the physical mechanisms of melt extraction, still yield a direct proportionality among these parameters. These relationships among P_o , $\bar{P}$, F_B and Z_c also satisfy common sense—higher temperatures lead to a larger melting interval, more melt and thicker ocean crust. This model also accounts well for the major-element and bathymetric data from normal ridges not influenced by hotspots⁷.

A partial exception to the correlation of all these parameters can occur with varying thicknesses of the lithospheric cap to the melting regime. For constant P_o , increasing lid thickness causes an increase in $\bar{P}$ but a decrease in extent of melting and crustal thickness, an effect that has been shown to be important for spreading rates less than 2 cm yr^{−1} (refs 6, 8). In this case as well, however, F_B and Z_c still correlate positively as the lid thickens.

These relationships do not take into account, however, the effect of water on mantle melting (Fig. 1). Water lowers the melting temperature of the mantle, and the addition of water to the source at constant potential temperature (T_p) causes melting to begin at a much higher pressure^{6,17}. But because water is strongly concentrated in the liquid and the amount of water is small, it is diluted by further melting higher in the melting column and has little effect on the maximum extent of melting^{18,19}. The net effect is to expand greatly the melting column by creating a large pressure interval at the base of the melting regime with very small amounts of melt present. The effects of this on F_B can be readily understood by considering that in a passive-flow, triangular melting regime, deepening the solidus by adding water increases the area of mantle flow that crosses the solidus, thereby increasing the denominator in the definition of F_B (ref. 15). The numerator is increased minimally because the productivity of deep hydrous melting is very small, hence the value of F_B decreases. Qualitatively, then, addition of water can reduce the mean extent of melting while leading to a small increase in crustal thickness.

Quantitative hydrous ridge models

We quantitatively demonstrate this effect with two models we have developed that add the effects of water to existing models of polybaric mid-ocean-ridge basalt melting and mixing. The model of ref. 7 (here called LKP), which is primarily based on olivine–liquid equilibria, has been modified by incorporating H cations as a simple diluent of other liquid species in the expression for the

olivine–liquid partition coefficient for Mg. This lowers the temperature of olivine–liquid equilibria linearly with liquid water content, and reproduces the experimental data of ref. 13. Here we call this model hLKP; for a more detailed reparameterization of wet melting with a similar approach, see ref. 20. The second model, pHMELTS, combines the thermodynamic melting model pMELTS²¹ with estimates of water solubility in nominally anhydrous minerals from the compilation in ref. 18; a detailed assessment of this model will be presented elsewhere (P.D.A., J. E. Dixon and C.H.L., manuscript in preparation). The effects of variations in potential temperature, water content, and other variables on the relationship between F_B and Z_c for both models are shown in Fig. 2. The model predictions differ in many details but the essential results are very similar. Most notably, for dry, passive-flow melting regimes with melting continuing to the base of the crust, increases in potential temperature increase both F_B and Z_c along concave-down curves, as expected. In both models the addition of water at constant potential temperature generates a steeply decreasing negative trend, such that F_B drops by a factor of two and crustal thickness increases by ~ 1 km between 0 and ~ 700 p.p.m. H_2O in the source (Fig. 2). Changing other parameters of melting does not alter the essential results: a variation from passive to active flow (keeping the depth to the solidus and the depth of the RMC equal) or an increase in the final pressure of melting both lead to positive correlation between F_B and Z_c . Addition of water, however, with other variables held constant, leads to a decrease in F_B for all these schemes, no matter whether melting is fractional, continuous or equilibrium.

The effect of water is counter-intuitive because it permits greater melt production at lower mean extents of melting. Because the concentrations of highly incompatible elements in the aggregate liquid are inversely proportional to the mean extent of melting,

increasing enrichments of incompatible elements relative to the source can be associated with increased melt production and crustal thickness.

Hotspot-affected ridges

These effects of water are able to account for a number of previously enigmatic aspects of oceanic basalt petrogenesis. One of the curious features of ‘hotspot’ magmatism is that hot regions should generate more melt and thus cause dilution of incompatible elements, and yet they invariably have exceptionally high concentrations of these elements. Looking at trace-element chemistry alone, one would be driven to infer very low extents of melting²², whereas increased crustal thickness and highly depleted residual peridotites imply higher extents of melting²³. Although isotopic evidence shows that the sources must be enriched, it is nonetheless hard to establish the quantitative magnitude of the enrichment. As hotspots are also high in water, we suggest that the bulk extents of melting are lower than would be inferred from the thick crust and shallow bathymetry, and hence the incompatible trace-element enrichments in the mantle sources are significantly lower than would have been inferred otherwise.

A simple illustration of this point is shown in Fig. 3a, where enrichment relative to the source is shown for Ba/Sm and Sm/Yb ratios for melting models that generate constant crustal thickness with varying water contents. As the water content of the source increases, the mean extent of melting decreases, and incompatible trace-element enrichments (for example, Ba/Sm) increase substantially, even if the source composition remains unchanged. Likewise, water addition raises the pressure of onset of melting and hence significantly increases the fraction of melt generated in the presence of residual garnet, even at constant crustal thickness, resulting in a

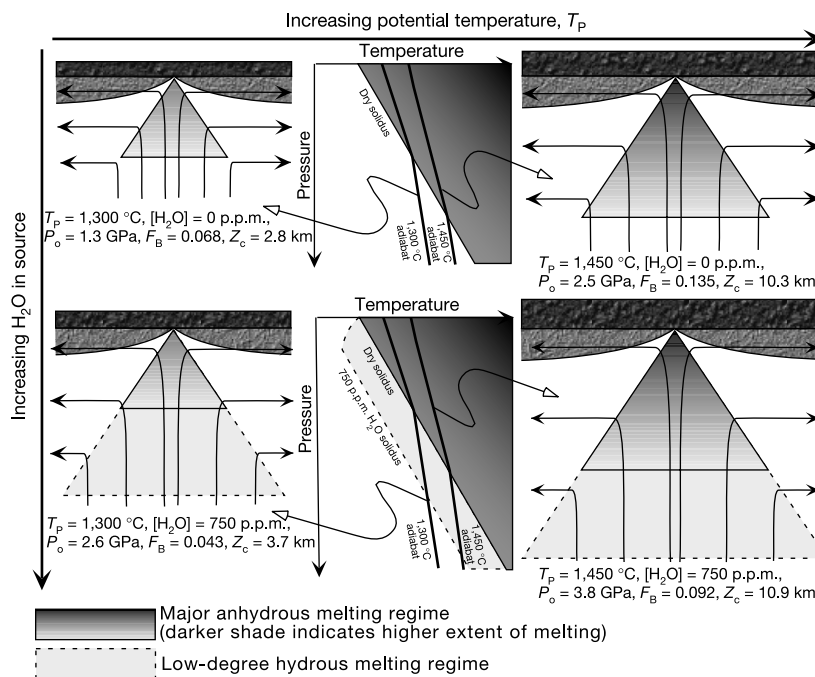


Figure 1 Conceptual models for the effects of potential temperature, T_p , and source H_2O concentration on mid-ocean-ridge melting regimes. Streamlines show a schematic passive corner flow in the asthenosphere. The shaded triangles indicate where melting is occurring. Variations in T_p control the depth at which adiabatically ascending mantle of a given composition crosses its solidus and begins melting. The presence of water induces an interval of low-productivity, low-degree melting below the level at which dry melting would begin. Increasing the water concentration increases the width of this interval, but high-productivity melting does not begin until

near the pressure and temperature of the dry solidus. The maximum extent of melting reached increases with the pressure interval between the nominal dry solidus and the base of the lithosphere, and is sensitive to T_p but not to water addition. The mean extent of melting, F_B , is given by the pressure at the base of the extracted crust divided by the pressure interval of melting¹⁵, and given the slopes of the wet and dry solidus curves and these productivity relations, F_B increases when extra crust is generated by an increase in T_p , but decreases when extra melting is driven by water addition. P_o , initial pressure of melting; Z_c , crustal thickness.

large increase in Sm/Yb. These coupled trace-element effects will be important in both ocean-ridge and ocean-island settings.

To explore the implications of these results in more detail, we examine data from two ocean ridges where depth and chemistry have been influenced by off-axis hotspots and are anomalous in a global context: the Galápagos spreading centre and the northern Mid-Atlantic Ridge near the Azores (Table 1). In both regions, high water contents are associated with shallow bathymetry^{23–25} and enrichment in incompatible elements and radiogenic isotope ratios^{23–25}. In the Galápagos, approaching longitude 91°W from either side, axial depth decreases from 2,800 m to ~1,700 m below sea level, the crustal thickness increases by ~2.5 km (refs 26, 27), and the fractionation-corrected water content of basalts reaches a peak with maximum values ~0.6 wt% (sources collected in the RIDGE petrological database²⁸). These effects have previously been attributed²⁶, in the context of an anhydrous model, to a 50 °C

increase in potential temperature; the anhydrous LKP or pMELTS models lead to similar results (Table 2). The Azores anomaly is similar but larger in magnitude: axial depth decreases from 3,500 to 1,200 m, crustal thickness increases by 3–4 km (ref. 29), and the fractionation and degassing-corrected water contents reach at least 0.9 wt% (ref. 12). Anhydrous melt production models

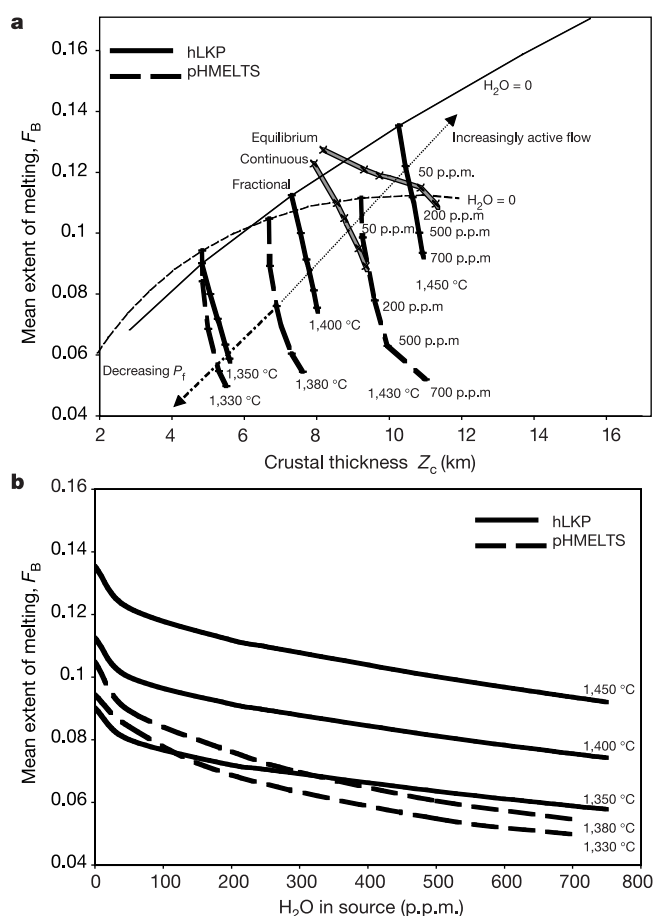


Figure 2 The relationships between mean extent of melting, F_B , crustal thickness, Z_c , and source H_2O content for mid-ocean-ridge melting regimes according to two models. The hLKP (continuous curves) and pMELTS (dashed curves) models described in the text account for addition of water to the mantle source of mid-ocean-ridge basalts. **a**, F_B versus Z_c . The fine curves labelled ' $H_2O = 0$ ' represent varying mantle T_P for dry sources in fractional fusion, passive-flow melting regimes^{2,7}. The bold curves with tick marks at 50, 200, 500 and 700 p.p.m. H_2O by weight show the effects of water addition to fractional, passive melting regimes at constant T_P . The effects of water addition are similar for continuous (1% residual porosity) and equilibrium melting (shown by grey curves). Changes to the parameters of the melting regimes such as increasing the pressure of final melting, P_f , due to conductive cooling (dot-dashed arrow towards origin) or a transition from passive flow towards pure active flow⁹ (dotted arrow away from origin) generate directly proportional changes in F_B and Z_c . **b**, The effect of source H_2O addition on F_B at constant T_P according to the hLKP (continuous curves) and pMELTS (dashed curves) models.

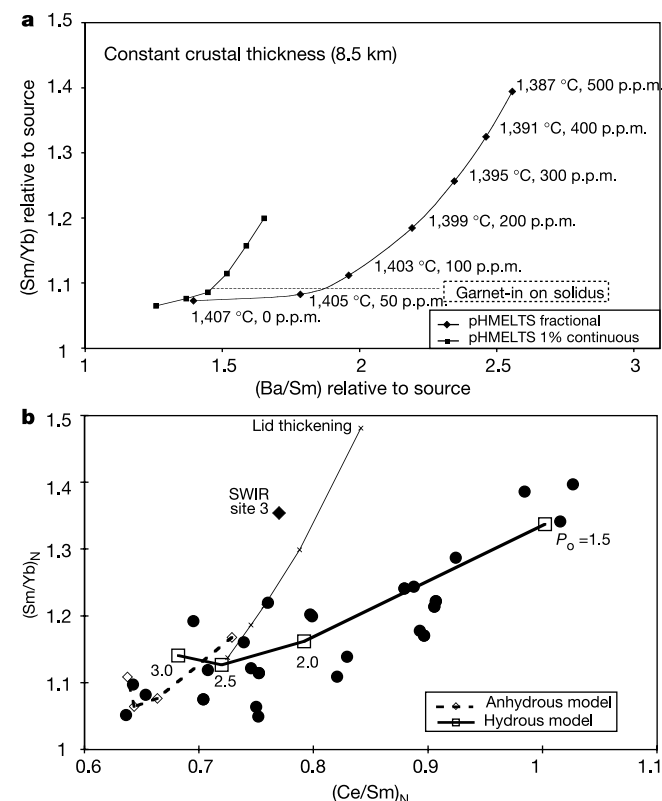


Figure 3 Trace-element systematics are strongly sensitive to the effects of water in the melting regime on both normal and hotspot-affected ridges. **a**, The presence of water changes the source region trace-element ratios that would be inferred from measured ratios in basalts and observations of Z_c . Results are shown for the pMELTS model with perfect fractional melting or continuous melting with 1% residual porosity. These calculations all yield 8.5 km of crust from a passive-flow melting regime; Z_c is held constant by lowering T_P as water content increases; points are labelled by T_P and source H_2O content. The drop in mean extent of melting associated with increasing water content causes an increase in (Ba/Sm) relative to the source. The increasing influence of garnet as water deepens the solidus is reflected in the increase in (Sm/Yb) relative to the source. **b**, Models of the rare-earth element ratios of isotopically normal MORB. Data points are segment-averaged, primitive-mantle-normalized (Sm/Yb)_N versus (Ce/Sm)_N for global ridge segments chosen with average $^{87}\text{Sr}/^{86}\text{Sr}$ between 0.7025 and 0.70265 to remove the effects of mantle heterogeneity, from the compilation in ref. 32. The global trend caused by mantle temperature variations in an anhydrous model with a homogenous source produces a very small range in this diagram. Addition of a constant 150 p.p.m. H_2O to the source, on the other hand, lowers F_B while increasing the ratio of maximum to minimum F_B , which gives the necessary range in (Ce/Sm)_N while increasing the contribution of melt from the garnet peridotite stability field and thereby raising (Sm/Yb)_N at the cold end. The point 'site 3' is from ref. 6, and lies above the global trend at a position best explained by imposition of a conductively cooled lid. Calculations for fractional melting are carried out by recalculating mineral assemblages every 1 kbar of pressure ascent using equations in ref. 7. Boxes on the 'hydrous model' curve are labelled by pressure of initial solidus intersection, P_o , in GPa. For anhydrous calculations, source concentrations of Ce, Sm and Yb are respectively 0.77, 0.32 and 0.45 p.p.m.; for hydrous calculations, 0.70, 0.31 and 0.44 p.p.m. are used. Partition coefficients for Ce, Sm and Yb are 0.086, 0.29 and 0.45 (clinopyroxene); 0.005, 0.02 and 0.11 (orthopyroxene); 0.008, 0.22 and 5 (garnet).

Table 1 Segment-average properties away from and near hotspot influence

	Long./lat.	Axial depth (m)	Crustal thickness (km)	Fe ₈	(H ₂ O) ₈	(La/Sm) _N
Galápagos	85–87°W	2,560	6	10.7	0.2	0.63
	90–92°W	1,650	8.5	9.6	0.45	1.37
Azores	33–35°N	2,870	6	10.0	0.18	0.76
	38–39.5°N	1,565	10	7.8	0.73	2.46

See text for details. Data are from refs 11, 12, 26, 28 and 29. Fe₈, total Fe as FeO corrected to 8 wt% MgO; (H₂O)₈, total dissolved H₂O corrected for degassing and fractionation to 8 wt% MgO; (La/Sm)_N, primitive-mantle-normalized ratio.

(Table 2) require a 75 °C temperature anomaly to generate the crustal thickness increase at the Azores platform. Concentrations and enrichment ratios of incompatible elements increase substantially as the depth shoals in both regions. The higher potential temperatures cause anhydrous models to predict increases of 10–30% in F_B at both the Galápagos and the Azores such that implied incompatible trace-element enrichments in the sources are significantly larger than the observed increases in the lavas. For example, treating water as a passive incompatible tracer, the implied water contents of the sources would reach ~400 p.p.m. for the Galápagos spreading centre and 650–800 p.p.m. for the Azores platform.

When instead we use hydrous melting models that produce the observed amount of water in the basalts and crustal thickness changes, we find that the Galápagos anomaly is explained by ≤300 p.p.m. H₂O in the source and an increase of ~40 °C in potential temperature, whereas the Azores anomaly is consistent with a peak of 500–600 p.p.m. H₂O in the source and a potential temperature anomaly of ~55 °C. In each case F_B drops by 40–60% approaching the most enriched regions, such that the source concentrations of incompatible elements need only increase by half as much as the observed increase in the enriched lavas (Table 2).

The importance of water to melting in these regions can be tested by investigating the effects on major-element compositions of basalts, which in these regions have substantially lower FeO at given MgO than basalts from equivalent depths elsewhere in the oceans. These low FeO values are not accounted for by the anhydrous mantle temperature model, as increasing potential temperature normally leads to deep melting and high FeO at given MgO (ref. 3) (Fig. 4). The effects of water on FeO can be evaluated by using the calculations above that constrain mantle temperature and extent of melting to calculate major-element compositions of mantle melts and then simulating fractionation of the melts for comparison to measured samples at equal MgO.

The calculations include the influence of water on the primary

melt composition, but the dominant effect of water is the suppression of plagioclase crystallization. This leads to low FeO* (total Fe expressed as FeO) in samples fractionated to 8% MgO (Fig. 4b)¹⁴, as expressed by the calculated parameter Fe₈. In the anhydrous calculations, the increases in potential temperature sufficient to generate the increased crustal thickness at the Galápagos or the Azores would lead to increases of 1–2% absolute in Fe₈—inconsistent with the data. On the other hand, using the hydrous hLKP or pHMELTS models, the increases in source water and smaller increases in potential temperature needed to match both crustal thickness and observed water contents lead to decreases of 1–2% absolute in Fe₈, consistent with the observations (Fig. 4a). Other major-element features of the data are also accounted for: the shallowest, most water-rich samples at both locations have high Al₈ and Ca₈ (ref. 11). Na₈ increases somewhat in the Galápagos, but is nearly constant over the Azores platform. High Al₈ and Ca₈ in multiply saturated samples are also signatures of enhanced crystallization of olivine (relative to plagioclase and clinopyroxene) owing to water. At the Azores, the combined effects of decreasing F_B and decreasing extent of fractionation to reach 8% MgO yield nearly constant Na₈ with no need for additional Na₂O in the source, whereas a dry model would require a significant Na₂O enrichment in the source to compensate for increasing extents of melting.

Si₈ is also low for the most water-rich samples. This reflects an increase in the mean pressure of melting, which is consistent with deep water-enhanced melting and/or an increase in potential temperature. Increased pressure is also indicated by a much stronger garnet influence on the trace-element patterns of the most water-rich lavas²⁵. Therefore addition of water clearly can lead to increased garnet influence and greater pressures of melting—at the same time as mean extents of melting are reduced (Fig. 3a).

The result that water addition causes a decrease in the mean extent of melting appears contrary to the results of ref. 30—a linear increase in extent of melting with addition of H₂O to the source of a suite of Mariana back-arc lavas. That study, however, examined

Table 2 Model results

	Long./lat.	Model	T_P , source H ₂ O	F_B	Fe ₈	Inferred source (La/Sm) _N
Galápagos	85–87°W	LKP dry	1,375 °C, 0	0.101	10.7	0.54
		hLKP wet	1,369 °C, 110 p.p.m.	0.082	10.6	0.49
		pMELTS dry	1,353 °C, 0	0.097	10.0	0.49
		pHMELTS wet	1,350 °C, 85 p.p.m.	0.065	10.5	0.44
Galápagos	90–92°W	LKP dry	1,421 °C, 0	0.122	11.7	1.23
		hLKP wet	1,412 °C, 375 p.p.m.	0.090	9.1	1.09
		pMELTS dry	1,407 °C, 0	0.107	11.5	1.09
		pHMELTS wet	1,395 °C, 300 p.p.m.	0.055	10.5	0.80
Azores	33–35°N	LKP dry	1,375 °C, 0	0.101	10.7	0.65
		hLKP wet	1,369 °C, 100 p.p.m.	0.082	10.6	0.59
		pMELTS dry	1,353 °C, 0	0.097	10.0	0.59
		pHMELTS wet	1,350 °C, 75 p.p.m.	0.065	10.5	0.53
Azores	38–39.5°N	LKP dry	1,447 °C, 0	0.134	11.9	2.24
		hLKP wet	1,436 °C, 625 p.p.m.	0.091	9.1	1.95
		pMELTS dry	1,432 °C, 0	0.109	12.3	1.95
		pHMELTS wet	1,405 °C, 500 p.p.m.	0.054	9.5	1.31

See text for details of models. T_P , potential temperature; F_B , mean extent of melting.

individual samples from a single ridge segment, and the consistency of that result with isothermal addition of H_2O to peridotite²¹ implies a small length scale of heterogeneity. Small water-rich heterogeneities are buffered to the temperature of surrounding peridotite, and therefore simply melt more at the same pressure as the adjoining anhydrous regions. Our results apply to pooled melts from the entire melting regime, where regional enrichments in source water content produce a large low extent of melting (F) tail deep in the melting regime. So although increased water in back-arc spreading centres may have a diagnostic chemical signature, it need not necessarily lead to larger mean extents of melting for the entire melting regime. For convergent margins, however, where there is no passive melting regime, increased water probably leads to both greater melt production and increased extents of melting. Understanding the effects of water in different tectonic settings requires combined examination of mantle flow patterns and the distribution and influence of water on melting.

Global consequences

The importance of the low- F tail produced by water is not limited to regions of enhanced water content such as the Azores and Galápagos. Even at the lower water contents of normal ocean ridges, water

will create a substantial low- F tail that extends deeper into the mantle. This will lower the mean F relative to anhydrous calculations. The effect is greater for colder mantle, because an equivalent low- F tail then is larger relative to the smaller, cold melting regime. This means that the ratio of minimum to maximum F_B for the global ocean-ridge system is greater when water is taken into account. The tail also allows a role for residual garnet in regions too cold for the anhydrous solidus to encounter garnet peridotite^{6,31}. Examination of global data for the rare-earth elements (REE) shows that both these effects are required for REE models to be consistent with major-element models for a homogeneous mantle source. Figure 3b shows REE data for normal ocean ridges from the RIDGE petrology database²⁸, screened to remove as much as possible the effects of mantle heterogeneity by choosing only ridge segments with mean $^{87}\text{Sr}/^{86}\text{Sr}$ from 0.7025 to 0.70265. There is no correlation within this subset between isotopes and trace-element enrichments. Although isotopic enrichment is associated with hotspots and thickened crust, in this isotopically filtered subset the highest light-REE enrichments are instead systematically from regions of thin crust. The anhydrous melting calculations are unable to produce the substantial range in either primitive-mantle-normalized $(\text{Ce}/\text{Sm})_N$ or $(\text{Sm}/\text{Yb})_N$ through variations in potential

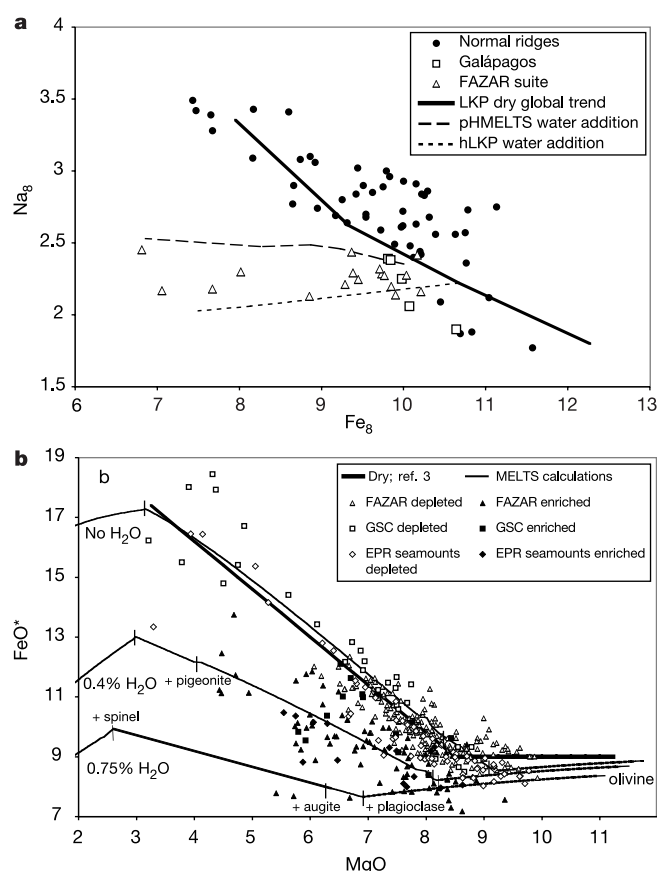


Figure 4 The influence of water on fractionation-corrected major-element composition of MORB. **a**, Segment-scale averages of FeO^* (total Fe expressed as FeO) and Na_2O concentrations in ocean-ridge basalts corrected for fractionation to 8% MgO , giving Fe_8 and Na_8 , respectively. The correction used the standard liquid line of descent given in ref. 3, which assumes that all mid-ocean-ridge basalts fractionate along a parallel trend in oxide versus MgO space with multiple saturation beginning at 8.5% MgO . The normal (that is, non-hotspot affected) ridge segments are well described by the global trend of T_F variations according to the dry LKP model⁷, but the Galápagos- and especially the Azores-affected segments have Fe_8 much too low for their Na_8 and Z_c values. However, addition of up to 750 p.p.m. water at constant T_F in either the hLKP or pHMELTS models causes apparent Fe_8 to decrease by ≥ 3 wt% at nearly

constant Na_8 (accompanied by a ~ 1 km increase in Z_c , see Fig. 2). **b**, The decrease in Fe_8 is dominated not by melting effects but by the effect of H_2O on low-pressure differentiation, where it both suppresses multiple saturation with plagioclase to lower MgO and decreases the slope of the multiply saturated liquid line of descent in FeO^* – MgO space¹⁴. The low FeO at any given MgO is a real effect generated by delay of multiple saturation during forward fractionation of primitive samples, whereas the low Fe_8 of water-rich samples in **a** is partly an artefact of correcting data back to 8% MgO using a liquid line of descent slope only appropriate for dry normal basalts. GSC is the Galápagos Spreading Centre; FAZAR is the Mid-Atlantic suite near the Azores¹¹. For this figure, the boundary between ‘depleted’ and ‘enriched’ is $\text{K}_2\text{O}/\text{TiO}_2 = 0.25$ for FAZAR, 0.11 for GSC, and 0.25 for East Pacific Rise (EPR) seamounts.

temperature⁵. On the other hand, varying mantle potential temperature with constant water content of ~150 p.p.m. accounts well for the global data—indeed it appears to be essential to produce the necessary ranges of, and correlation between, crustal thickness and (Ce/Sm)_N (as well as Na₈). An important implication is that the mantle source of normal MORB is more depleted (~10% lower in (Ce/Sm)_N) than would be inferred from anhydrous batch melting calculations.

A trend similar to the data in Fig. 3b can also be obtained from models where variable source heterogeneity is imposed by a source component carrying both water and incompatible elements at nearly constant potential temperature (such as we invoke above for hotspot-affected regions). To explain the isotopically normal ridge segments, however, this heterogeneity would have to be very young, and this model also would not account for the significant differences in crustal thickness. Note also that the effects of lithospheric lid thickening caused by surface cooling can be distinguished from mantle temperature variations or addition of low-*F* melt component by relatively high (Sm/Yb)_N ratios, as observed for some, but not all, segments along the slow-spreading southwest Indian ridge⁶, and the Gakkel ridge.

Finally, the reasoning behind the large effect of the low-*F* tail on cold mantle melting regimes can also be applied to hotspots that rise beneath thick lithosphere. A short melting column caused by the truncation of the melting regime by the lithosphere will also be strongly influenced by a long, low-*F* tail of melt. This effect could cause actual extents of melting to be a factor of two lower than inferred from anhydrous models, and sources to be correspondingly less enriched. □

Received 20 November 2002; accepted 14 January 2003; doi:10.1038/nature01429.

- Dick, H. J. B., Fisher, R. L. & Bryan, W. B. Mineralogic variability of the uppermost mantle along mid-ocean ridges. *Earth Planet. Sci. Lett.* **69**, 88–106 (1984).
- Asimow, P. D., Hirschmann, M. M. & Stolper, E. M. Calculation of peridotite partial melting from thermodynamic models of minerals and melts. IV. Adiabatic decompression and the composition and mean properties of mid-ocean ridge basalts. *J. Petrol.* **42**, 963–998 (2001).
- Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
- McKenzie, D. P. & Bickle, M. J. The volume and composition of melt generated by extension of the lithosphere. *J. Petrol.* **29**, 625–679 (1988).
- Shen, Y. & Forsyth, D. W. Geochemical constraints on initial and final depths of melting beneath mid-ocean ridges. *J. Geophys. Res.* **100**, 2211–2237 (1995).
- Robinson, C. J., Bickle, M. J., Minshull, T. A., White, R. S. & Nichols, A. R. L. Low degree melting under the Southwest Indian Ridge: the roles of mantle temperature, conductive cooling and wet melting. *Earth Planet. Sci. Lett.* **188**, 383–398 (2001).
- Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 183–280 (American Geophysical Union, Washington DC, 1992).
- White, R. S., Minshull, T. A., Bickle, M. J. & Robinson, C. J. Melt generation at very slow-spreading oceanic ridges: Constraints from geochemical and geophysical data. *J. Petrol.* **42**, 1171–1196 (2001).
- Forsyth, D. W. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 1–66 (American Geophysical Union, Washington DC, 1992).
- Johnson, K. T. M., Dick, H. J. B. & Shimizu, N. Melting in the oceanic upper mantle: An ion microprobe study of diopsides in abyssal peridotites. *J. Geophys. Res.* **95**, 2661–2678 (1990).
- Asimow, P. D. & Langmuir, C. H. Segment-scale and regional systematics from 33°N to 41°N on the Mid-Atlantic Ridge: Results from the FAZAR Cruise. *Eos* **79**, F938–F939 (1998).
- Dixon, J. E., Leist, L., Langmuir, C. H. & Schilling, J.-G. Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* **420**, 385–389 (2002).
- Gaetani, G. A. & Grove, T. L. The influence of water on melting of mantle peridotite. *Contrib. Mineral. Petrol.* **131**, 323–346 (1998).
- Michael, P. J. & Chase, R. L. The influence of primary magma composition H₂O and pressure on mid-ocean ridge basalt differentiation. *Contrib. Mineral. Petrol.* **96**, 245–263 (1987).
- Plank, T., Spiegelman, M., Langmuir, C. H. & Forsyth, D. W. The meaning of “mean *F*”: Clarifying the mean extent of melting at ocean ridges. *J. Geophys. Res.* **100**, 15045–15052 (1995).
- Plank, T. & Langmuir, C. H. Effects of the melting regime on the composition of oceanic crust. *J. Geophys. Res.* **97**, 19749–19770 (1992).
- McKenzie, D. The extraction of magma from the crust and mantle. *Earth Planet. Sci. Lett.* **74**, 81–91 (1985).
- Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
- Hirschmann, M. M., Asimow, P. D., Ghiorso, M. S. & Stolper, E. M. Calculation of peridotite partial melting from thermodynamic models of minerals and melts III. Controls on isobaric melt production and the effect of water on melt production. *J. Petrol.* **40**, 831–851 (1999).
- Katz, R. F., Spiegelman, M. & Langmuir, C. H. Models of equilibrium and reactive melting in a subduction zone setting. *Geochem. Geophys. Geosyst.* (submitted).
- Ghiorso, M. S., Hirschmann, M. M., Reiners, P. W. & Kress, V. C. The pMELTS: A revision of MELTS for improved calculation of phase relations and major element partitioning related to partial melting of the mantle to 3 GPa. *Geochem. Geophys. Geosyst.* **3**, 10.1029/2001GC000217 (2002).
- Kay, R. W. & Gast, P. W. The rare earth content and origin of alkali-rich basalts. *J. Geol.* **81**, 653–682 (1973).
- Bonatti, E. Not so hot “hot spots” in the oceanic mantle. *Science* **250**, 107–111 (1990).
- Schilling, J.-G., Bergeron, M. B. & Evans, R. Halogens in the mantle beneath the North Atlantic. *Phil. Trans. R. Soc. Lond. A* **297**, 147–178 (1980).
- Schilling, J.-G. Azores mantle blob: the rare earth evidence. *Earth Planet. Sci. Lett.* **25**, 103–115 (1975).
- Ito, G. T. & Lin, J. Mantle temperature anomalies along the present and paleoaxes of the Galápagos spreading center as inferred from gravity analyses. *J. Geophys. Res.* **100**, 3733–3745 (1995).
- Detrick, R. S. *et al.* Correlated geophysical, geochemical, and volcanological manifestations of plume-ridge interaction along the Galápagos Spreading Center. *Geochem. Geophys. Geosyst.* **3**, 10.1029/2002GC000350 (2002).
- Lehnert, K., Su, Y., Langmuir, C. H., Sarbas, B. & Nohl, U. A global geochemical database structure for rocks. *Geochem. Geophys. Geosyst.* **1**, 10.1029/1999GC000026 (2000).
- Ghiorso, M. S., Needham, H. D. & Renard, V. Gravity anomalies and crustal thickness variations along the Mid-Atlantic Ridge between 33°N and 40°N. *J. Geophys. Res.* **100**, 3767–3787 (1995).
- Stolper, E. M. & Newman, S. The role of water in the petrogenesis of Mariana trough magmas. *Earth Planet. Sci. Lett.* **121**, 293–325 (1994).
- Langmuir, C. H. Deep low *F* melts and the global systematics of MORB. *Eos* **76**, F694–F695 (1995).
- Su, Y.-J. *Database Applications to the Petrogenesis of Ocean Ridge Basalts*. Thesis, Columbia Univ. (2002).

Acknowledgements We thank P. Michael for comments and suggestions. This work was supported by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.D.A. (e-mail: asimow@gps.caltech.edu).

Investigation of the obscuring circumnuclear torus in the active galaxy Mrk231

Hans-Rainer Klöckner^{*†}, Willem A. Baan[†] & Michael A. Garrett[‡]

^{*} Kapteyn Institute, PO Box 800, 9700 AV Groningen, The Netherlands

[†] ASTRON, Westerbork Observatory, [‡] JIVE, Joint Institute for VLBI in Europe, PO Box 2, 7990 AA Dwingeloo, The Netherlands

Active galaxies are characterized by prominent emission from their nuclei. In the 'unified' view of active galaxies, the accretion of material onto a massive compact object—now generally believed to be a black hole—provides the fundamental power source¹. Obscuring material along the line of sight can account for the observed differences in the nuclear emission^{2,3}, which determine the classification of AGN (for example, as Seyfert 1 or Seyfert 2 galaxies). Although the physical processes of accretion have been confirmed observationally^{4,5}, the structure and extent of the obscuring material have not been determined. Here we report observations of powerful hydroxyl (OH) line emissions that trace this obscuring material within the circumnuclear environment of the galaxy Markarian 231. The hydroxyl (mega)-maser emission shows the characteristics of a rotating, dusty, molecular torus (or thick disk) located between 30 and 100 pc from the central engine. We now have a clear view of the physical conditions, the kinematics and the spatial structure of

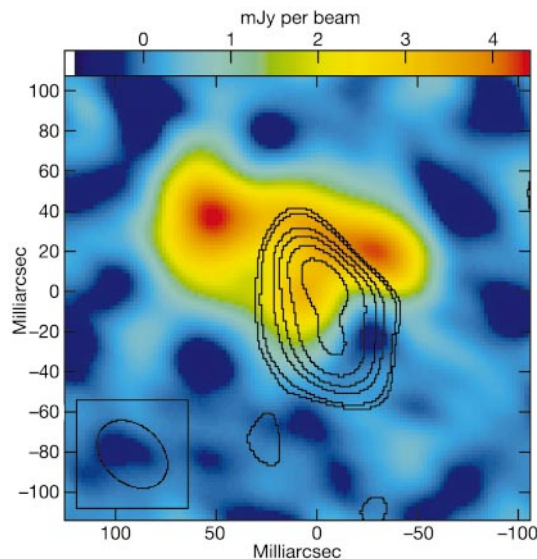


Figure 1 Hydroxyl and radio-continuum emission in the nucleus of Mrk231. The integrated OH-line emission (pseudo-colour, in mJy per beam) is superimposed on the nuclear continuum emission (contours, peak 39 mJy per beam). The contour levels are a geometric progression in factors of $\sqrt{2}$ starting at 4 mJy per beam. These EVN observations of Mrk231 were made on September 1999 for 12 h in dual-polarization phase-referencing mode at 1,599 MHz, and with 256 spectral channels across a total bandwidth of 8 MHz. The synthesized beam is 39 mas $\times$ 28 mas, where 1 mas corresponds to a size of 0.83 pc at the distance of Mrk231 (172 Mpc, assuming $q_0 = 0.5 \text{ km s}^{-1}$ and $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$). The data were correlated at the EVN correlator at the Joint Institute for VLBI in Europe (JIVE), and later transferred to the AIPS packages for calibration of the complex visibilities. The integrated line emission image was obtained by averaging over the velocity range of the OH lines, after subtraction of the radio continuum, obtained by averaging the channels without line emission.

this material on intermediate size scales, confirming the main tenets of unification models.

Observations of powerful molecular water-vapour and hydroxyl megamaser emissions, which are a million times more luminous than their galactic counterparts⁶, have provided unique information about distinct regions within galactic nuclei. Water-vapour masers can map out the central few parsecs of an active galactic nucleus, tracing the rotation of nuclear accretion disks and the state of molecular material surrounding the radio jets⁷. The inferred geometry and thickness of these disks, however, cannot account for the observed obscuration signature and the radiation patterns seen in active galactic nuclei (AGN)^{7,8}. On the other hand, the hydroxyl masers in galactic nuclei are spread across hundreds of parsecs^{9–11}. Their tight correlation with the infrared emission¹² indicates a strong association with the dusty obscuring material along the line of sight towards the nucleus^{2,3}. To study the distribution of the hydroxyl megamaser emission with respect to the nuclear engine at a resolution of tens of parsecs in Mrk231, we performed very-long-baseline interferometry (VLBI) radio observations with the European VLBI Network (EVN). Mrk231 is the most luminous infrared galaxy in the local universe (distance of 172 Mpc), and is undergoing a merger as indicated by the high dust content and the prominent tidal tails observed at optical wavelengths. The nuclear power source has been classified¹³ as a Seyfert 1, which in the

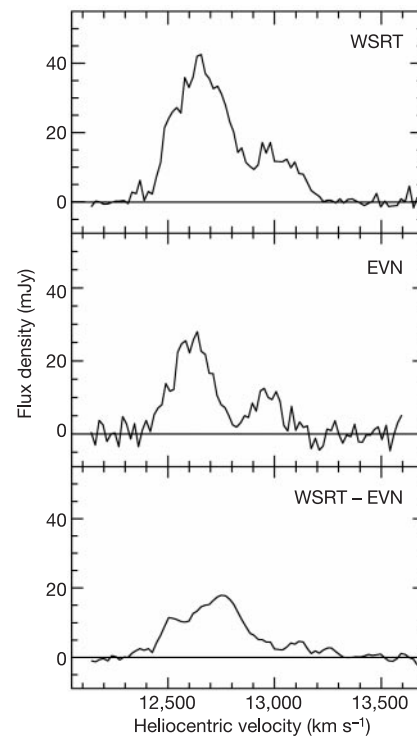


Figure 2 Hydroxyl emission spectra of Mrk231 traced at different scale sizes. Top panel, the WSRT spectrum taken at 14-arcsec resolution; middle panel, the spectrum for the EVN observations at 39-mas resolution; bottom panel, the difference spectrum of both observations. The emission line features show both OH main lines at 1,667 and 1,665 MHz that are offset by 365 km s^{-1} ; the velocity scale refers to the 1,667 MHz emission line by assuming a heliocentric velocity of $V_{\text{centre}} = 12,650 \text{ km s}^{-1}$ based on the CO(1–0) centre velocity¹⁹. The spectral resolution is 18 km s^{-1} per channel. The OH linewidth in the EVN data is $214 \pm 11 \text{ km s}^{-1}$ (FWHM), which is 70 km s^{-1} less than for the WSRT data. The residual emission line in the bottom frame counts for 44% of the OH emission in Mrk231, and is systematically red-shifted by around 45 km s^{-1} . Similar flux discrepancies, but with no distinct velocity offset, have been found for other OH-MM galaxies studied at high resolution^{9–11}. Similar velocity offsets have also been found for the CO(2–1) emission¹⁴ and the H₁ absorption in Mrk231 (ref. 20).

unification scheme would suggest that the observer has a direct, and almost unobscured, line-of-sight view into the broad-line emission region at parsec scales. The location of the nuclear engine on similar scale sizes is traced by the radio outflow¹⁴, which extends to almost a hundred parsecs¹⁵.

The EVN data reveal an OH emission structure with an east–west extent of 130 pc that is straddling a north–south elongated radio continuum emission (Fig. 1). This continuum emission has been further resolved into a lobe–core–lobe structure¹⁵ of 80 pc at position angle $PA = 8^\circ$, which is found to be misaligned by 57° with the core radio structure on parsec scales¹⁴. The emission spectrum integrated across the whole OH emitting region displays the two main hydroxyl lines at 1,667 and 1,665 MHz showing a line-width of around 200 km s^{-1} (full-width at half-maximum, FWHM) at a centroid velocity of $12,610 \text{ km s}^{-1}$ (Fig. 2). The EVN data and also the intermediate-scale MERLIN¹⁶ observations account for only half of the integrated line emission seen by the Westerbork Synthesis Radio Telescope (WSRT; see Fig. 2) and previous single-dish measurements⁶. This missing emission component must therefore originate within the resolution gap between the EVN and WSRT observations, in an extended region between 190 pc and 11 kpc of the nucleus. This component is probably associated with the nuclear continuum emission¹⁷ and the CO(1–0) line emission^{18,19}, both of which are located within the inner kiloparsec of Mrk231. Figure 3a presents the spatial velocity distribution, which is similar for each of the two hydroxyl lines and which has an overall velocity shift of about 160 km s^{-1} from the northwest towards the southeast. The velocity width of the emission lines has an average value of about 70 km s^{-1} (FWHM) across the northern part of the radio continuum emission. On the other hand, there is also a distinct OH emission region at the northwestern edge of the continuum emission that has a velocity similar to that seen towards the centre of the radio core, and which has a distinctly higher velocity width of 85 km s^{-1} (Fig. 3b).

The nature of the OH emission observed in the nucleus of Mrk231 adds a new component to the hierarchical structure of this system, and complements the information obtained with other tracers at lower resolution. We find that the OH velocity field with

its symmetry axis at $PA = 34^\circ$ does not agree with that of the kiloparsec-scale CO(1–0) structure^{18,19} having a $PA = 0^\circ$. Neither does it agree with the large-scale radio outflow ($PA = 8^\circ$), or the more compact CO(2–1) distribution¹⁹ extending up to 850 pc at $PA = 18^\circ$. However, the OH velocity field is in agreement with that of the line-of-sight H I absorption²⁰ at $PA = 27^\circ$. The H I absorption is detected against a diffuse radio-emitting halo with an extent of about 350 pc and an orientation of $PA = 22^\circ$. This diffuse radio halo, which is possibly associated with a circumnuclear starburst¹⁷, could then also trace the dense and opaque interstellar medium of a disk or torus²⁰ inclined at 56° . The integrated H I velocity dispersion²⁰ of $193 \pm 25 \text{ km s}^{-1}$ agrees well with that of the hydroxyl lines (Fig. 2). This suggests that the OH emission indeed originates in the central regions of a circumnuclear rotating disk or torus having an outer structure that is traced by the CO(2–1) emission.

The energy source responsible for the excitation of the OH molecules is the extreme far-infrared radiation¹² field in Mrk231. The observed far-infrared spectral temperature of around 43 K provides ideal pumping conditions for the OH clouds, which could range in size from 10 pc up to 100 pc (refs 21, 22). The ratio of the intensity of the two maser lines varies rather smoothly across the entire emission structure and has an average value of 1.8, as expected for local temperature equilibrium. This would indicate that the observed emission lines are optically thin and are mostly unsaturated masers. Assuming that the diffuse radio continuum, which serves as a background for the nuclear H I absorption, also serves as a background for the observed OH masers, this emission could be explained in terms of exponential amplification with a maximal gain of about 2.2 within classical OH maser models^{6,23}. However, extreme values for the line ratio have been found in the discrepant northwestern region discussed above. This region must have a completely different pumping environment than the rest of the structure. The unusual blue-shifted velocity, the slightly higher velocity dispersion, and the line ratios of this discrepant region are all consistent with the theoretical predictions^{6,22,24} for an interaction region between the radio outflow and the molecular environment.

The molecular environment described above for the nuclear region of Mrk231 is very similar to the dust structures found in

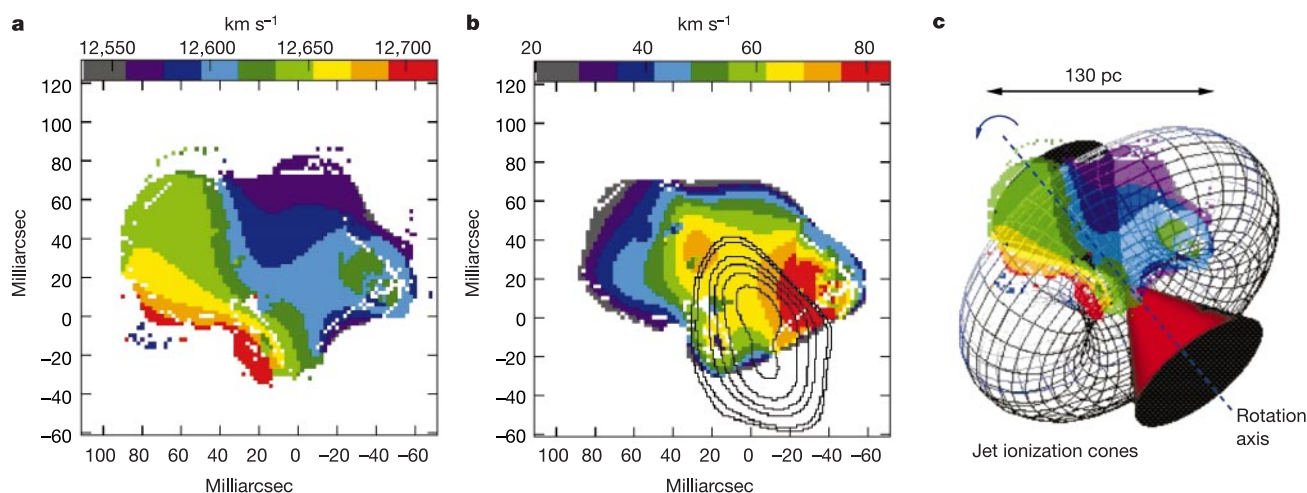


Figure 3 Circumnuclear kinematics in Mrk231. **a**, The velocity field of the OH 1,667 MHz emission in the EVN data is shown in pseudo-colours. The spatial resolution is the same as for Fig. 1, and the velocity scale is the same as in Fig. 2. The overall velocity pattern reveals an east–west velocity gradient of 1.41 km s^{-1} per parsec over the entire OH emission such that the northwestern edge of the torus moves towards the observer. **b**, The linewidth of the 1,667 MHz emission lines (in km s^{-1}) is superimposed on the nuclear radio continuum emission in contours displayed in a similar way as in Fig. 1. The diagram shows that the velocity dispersion varies significantly across the emission region. The

velocity dispersion structure shows several distinct regions, but with the highest values to be found at the western edge of the radio continuum source. **c**, The inferred model of the nuclear torus in Mrk231 is displayed as a wire diagram with symmetric ionization cones. This model takes into account all large-scale characteristics of the nuclear radio emission and the OH emission. The symmetry axis of the torus has been inclined upwards by 56° , and then rotated anticlockwise by 35° . The molecular material moves from top right to bottom left (northwest to southeast). Virial estimates of the central mass concentration give $(7.2 \pm 3.8) \times 10^7$ solar masses.

active galaxies such as NGC4261 using the Hubble Space Telescope²⁵. We have applied an axisymmetric torus model with a circular cross-section to the EVN data, in order to determine the structural boundaries in the nuclear region. The compact molecular structure surrounding the nucleus is misaligned with the large-scale molecular disk of the galaxy by 34°. The radius of the inner cavity of the torus would be about 30 pc, as suggested by the location of the northwestern interaction region. The outer edge of the torus spans a diameter of 200 pc, which is based on the 4-mJy contours of the diffuse radio emission²⁰ and on far-infrared blackbody estimates¹⁹. This region also serves as a radio background for the OH amplification process. These parameters result in a torus structure with a radius of 65 pc and a thickness of 70 pc, which accounts for an obscuration angle of 60° centred on the plane of the torus and which is representative of values found for other galaxies²⁶. The orientation and the shape of the extended radio contours and the symmetry axis of the velocity field indicate that the torus is tilted upward from the line of sight by 56°, and is rotated anticlockwise by 35°. This inferred model for the torus is shown as a wire-structure representation in Fig. 3c. The nuclear radiation cones with opening angles of 60°, assumed to be similar to those seen²⁷ in M87, represent the directions with an unobscured view of the nucleus. The radio outflow follows a twisted path within these cones, starting at the nucleus (at PA = 65°) and ending up as a double source aligned with the symmetry axis of the outer disk (at PA = 8°)^{14,15}. As a result, this parameterization indicates that the nuclear accretion disk, which collimates the jet outflow inside the central cavity, is actually misaligned by 30° with the plane of the molecular torus.

In reality, the proposed torus structure does not need to have a circular cross-section. In addition, it will most probably be wrapped inside a cocoon-like surface region with a higher temperature, and will have the same outer extent of 460 pc as the diffuse radio emission and the H I absorbing gas²⁰. However, the representation of the inner part of this rotating, dusty and molecular torus in Mrk231 is consistent with (all) current models of galactic nuclei and theoretical investigations³, and supports the unification schemes for AGN. Whether a different orientation of the torus and the nuclear accretion disk represents a special case or is a general characteristic of active nuclei needs to be investigated with VLBI observations of other megamaser galaxies hosting a similar nuclear power source. □

Received 22 July; accepted 17 December 2002; doi:10.1038/nature01381.

- Rees, M. J. Black hole models for active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
- Krolik, J. H. & Begelman, M. C. Molecular tori in Seyfert galaxies: Feeding the monster and hiding it. *Astrophys. J.* **329**, 702–711 (1988).
- Krolik, J. H. *Active Galactic Nuclei* (Princeton Series in Astrophysics, Princeton Univ. Press, Princeton, New Jersey, 1999).
- Greenhill, L. J. *et al.* Detection of a subparsec diameter disk in the nucleus of NGC 4258. *Astrophys. J.* **440**, 619–627 (1995).
- Gallimore, J. F., Baum, S. F. & O'Dea, C. P. A direct image of the obscuring disk surrounding an active galactic nucleus. *Nature* **388**, 852–854 (1997).
- Baan, W. A. Powerful extragalactic masers. *Nature* **315**, 26–31 (1985).
- Maloney, P. R. Powerful water masers in active galactic nuclei. *Publ. Astron. Soc. Pacif.* **19**, 401–421 (2002).
- Herrnstein, J. R. *et al.* Polarimetric observations of the masers in NGC 4258: an upper limit on the large-scale magnetic field 0.2 parsecs from the central engine. *Astrophys. J.* **508**, 243–247 (1998).
- Diamond, P. J., Lonsdale, Co. J., Lonsdale, Ca. J. & Smith, H. E. Global VLBI observations of the compact OH megamaser emission from III Zw 35 and IRAS 17208-0014. *Astrophys. J.* **511**, 178–184 (1999).
- Lonsdale, Co. J., Lonsdale, Ca. J., Diamond, P. J. & Smith, H. E. Symmetric parsec-scale OH megamaser structures in Arp 220. *Astrophys. J.* **507**, 615–654 (1998).
- Pihlström, Y. M., Conway, J. E., Booth, R. S., Diamond, P. J. & Polatidis, A. G. EVN and MERLIN observation of III Zw 35. *Astron. Astrophys.* **377**, 377–413 (2001).
- Baan, W. A. Infrared properties of OH galaxies. *Astrophys. J.* **338**, 804–811 (1989).
- Baan, W. A., Salzer, J. J. & LeWinter, D. Optical classification of megamaser galaxies. *Astrophys. J.* **509**, 633–645 (1998).
- Ulvestad, J. S. *et al.* Subrelativistic radio jets and parsec-scale absorption in two Seyfert galaxies. *Astrophys. J.* **517**, L81–L84 (1999).
- Ulvestad, J. S., Wrobel, J. M. & Carilli, C. L. Radio continuum evidence for outflow and absorption in the Seyfert 1 galaxy Markarian 231. *Astrophys. J.* **516**, 127–140 (1998).
- Richards, A. *et al.* in *Galaxies and their Constituents at Highest Angular Resolutions* (eds Schilizzi, R. T.,

- Vogel, S. F. P. & Elvis, M. S.) 212–215 (IAU Symposium 205, Astronomical Society of the Pacific, San Francisco, 2001).
- Taylor, G. B., Silver, C. S., Ulvestad, J. S. & Carilli, C. L. The starburst in the central kiloparsec of Markarian 231. *Astrophys. J.* **519**, 185–190 (1999).
- Bryant, P. M. & Scoville, N. Z. High-resolution CO observations of the ultraluminous infrared galaxy Markarian 231. *Astrophys. J.* **457**, 678–692 (1996).
- Downes, D. & Solomon, P. M. Rotating nuclear rings and extreme starbursts in ultraluminous galaxies. *Astrophys. J.* **507**, 615–654 (1998).
- Carilli, C. L., Wrobel, J. M. & Ulvestad, J. S. A subkiloparsec disk in Markarian 231. *Astron. J.* **115**, 928–937 (1998).
- Baan, W. A., Henkel, C. & Güsten, R. Rotationally excited OH in megamaser galaxies. *Astron. Astrophys.* **185**, 14–24 (1987).
- Randell, J., Field, D., Jones, K. N., Yates, J. A. & Gray, M. D. The OH zone in OH megamaser galaxies. *Astron. Astrophys.* **300**, 659–674 (1995).
- Baan, W. A., Wood, P. A. D. & Haschick, A. D. Broad hydroxyl emission in IC 4553. *Astrophys. J.* **260**, L49–L52 (1982).
- Keiichi, W. & Norman, C. A. Obscuring material around Seyfert nuclei with starbursts. *Astrophys. J.* **566**, L21–L24 (2002).
- v. Langevelde, H. *et al.* A thin circumnuclear disk in NGC 4261. *Astron. Astrophys.* **354**, L45–L48 (2000).
- Schmitt, H. R. *et al.* Testing the unified model with a infrared-selected sample of Seyfert galaxies. *Astrophys. J.* **555**, 663–672 (2001).
- William, J., Biretta, J. A. & Livio, M. Formation of the radio jet in M87 at 100 Schwarzschild radii from the central black hole. *Nature* **401**, 891–892 (1999).

Acknowledgements We thank C. Carilli for providing a map of the diffuse continuum structure in Mrk231. H.-R.K. thanks O. Möller for advice on programming in OpenGL software. The European VLBI Network is a joint facility of European, Chinese, South African and other radio astronomy institutes funded by their national research councils. The Westerbork Synthesis Radio Telescope is operated by ASTRON (Netherlands Foundation for Research in Astronomy) with support from the Netherlands Foundation for Scientific Research (NWO).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.-R.K. (e-mail: hrkloek@astro.rug.nl).

Quantum oscillations in two coupled charge qubits

Yu. A. Pashkin^{*,†}, T. Yamamoto^{*,†}, O. Astafiev^{*}, Y. Nakamura^{*,†}, D. V. Averin[§] & J. S. Tsai^{*,†}

^{*} The Institute of Physical and Chemical Research (RIKEN), Wako, Saitama 351-0198, Japan

[†] NEC Fundamental Research Laboratories, Tsukuba, Ibaraki 305-8501, Japan

[§] Department of Physics and Astronomy, SUNY Stony Brook, New York 11794-3800, USA

A practical quantum computer¹, if built, would consist of a set of coupled two-level quantum systems (qubits). Among the variety of qubits implemented², solid-state qubits are of particular interest because of their potential suitability for integrated devices. A variety of qubits based on Josephson junctions^{3,4} have been implemented^{5–8}; these exploit the coherence of Cooper-pair tunnelling in the superconducting state^{5–10}. Despite apparent progress in the implementation of individual solid-state qubits, there have been no experimental reports of multiple qubit gates—a basic requirement for building a real quantum computer. Here we demonstrate a Josephson circuit consisting of two coupled charge qubits. Using a pulse technique, we coherently mix quantum states and observe quantum oscillations, the spectrum of which reflects interaction between the qubits. Our results demonstrate the feasibility of coupling multiple solid-state qubits, and indicate the existence of entangled two-qubit states.

One of the physical realizations of a solid-state qubit is provided

[†] Permanent address: Lebedev Physical Institute, Moscow 117924, Russia.

by a Cooper-pair box¹¹. The two charge states of the box, say $|0\rangle$ and $|1\rangle$, differing by one Cooper pair are coherently mixed by the Josephson coupling, as has been confirmed experimentally^{12,13}. Quantum state manipulation of such a system can be done by using a non-adiabatic pulse technique, and read-out can be performed by a properly biased probe electrode⁵. Here we take one step further on the way to implementation of quantum logic gates by integrating two charge qubits and demonstrating their interaction.

The two charge qubits of our circuit are electrostatically coupled by an on-chip capacitor (Fig. 1). The right qubit has a SQUID (superconducting quantum interference device) geometry to allow control of the Josephson coupling to its reservoir. Both qubits have a common pulse gate but separate d.c. gates, probes and reservoirs. The pulse gate has nominally equal coupling to each box. The hamiltonian of the system in the two-qubit charge basis $|00\rangle$, $|10\rangle$, $|01\rangle$ and $|11\rangle$ reads:

$$H = \begin{bmatrix} E_{00} & -\frac{1}{2}E_{J1} & -\frac{1}{2}E_{J2} & 0 \\ -\frac{1}{2}E_{J1} & E_{10} & 0 & -\frac{1}{2}E_{J2} \\ -\frac{1}{2}E_{J2} & 0 & E_{01} & -\frac{1}{2}E_{J1} \\ 0 & -\frac{1}{2}E_{J2} & -\frac{1}{2}E_{J1} & E_{11} \end{bmatrix} \quad (1)$$

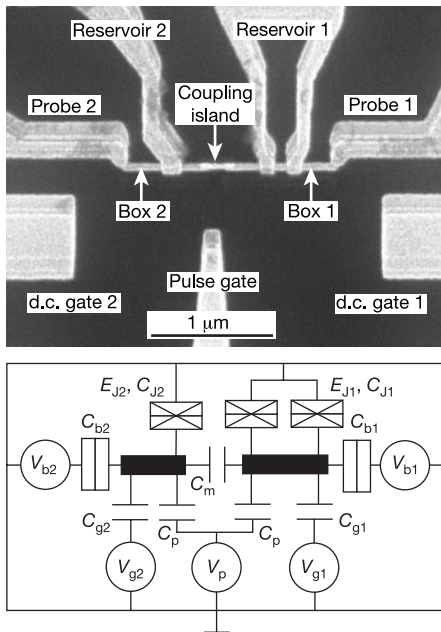


Figure 1 Two capacitively coupled charge qubits. **a**, Scanning electron micrograph of the sample. The qubits were fabricated by electron-beam lithography and three-angle evaporation of Al (light areas) on a SiN_x insulating layer (dark) (see ref. 5 for fabrication details). Two qubits are coupled by an additional coupling island overlapping both Cooper-pair boxes. Although the coupling island has a finite tunnelling resistance (~ 10 M Ω) to the boxes, we consider the coupling as purely capacitive (represented by a single capacitor in the equivalent circuit) because all the tunnelling processes are completely blocked. The estimated capacitance of the island to the ground is ~ 1 aF. **b**, Equivalent circuit of the device. The parameters (see text for definitions) obtained from the d.c. measurements are: $C_{J1} = 620$ aF, $C_{J2} = 460$ aF, $C_{b1} = 41$ aF, $C_{b2} = 50$ aF, $C_{g1} = 0.60$ aF, $C_{g2} = 0.61$ aF, $C_p \approx 1$ aF and $C_m = 34$ aF, and the corresponding energies are $E_{c1} = 484$ μ eV (117 GHz in frequency units), $E_{c2} = 628$ μ eV (152 GHz) and $E_m = 65$ μ eV (15.7 GHz). Josephson coupling energies, $E_{J1} = 55$ μ eV (13.4 GHz) and $E_{J2} = 38$ μ eV (9.1 GHz), were determined from the single qubit measurements described in the text. Probe junction tunnel resistance is equal to 31.6 M Ω (left) and 34.5 M Ω (right). Superconducting energy gap is 210 μ eV. Black bars denote Cooper-pair boxes. Vertical rectangles containing a single line represent tunnel junctions without Josephson coupling; horizontal, crossed rectangles represent Josephson tunnel junctions.

where $E_{n_1 n_2} = E_{c1}(n_{g1} - n_1)^2 + E_{c2}(n_{g2} - n_2)^2 + E_m(n_{g1} - n_1) \times (n_{g2} - n_2)$ is the total electrostatic energy of the system (n_1 , $n_2 = 0, 1$ is the number of excess Cooper pairs in the first and the second box), $E_{J1}(E_{J2})$ is the Josephson coupling energy of the first (second) box and the reservoir, $E_{c1,2} = 4e^2 C_{\Sigma 2,1} / 2(C_{\Sigma 1} C_{\Sigma 2} - C_m^2)$ are the effective Cooper-pair charging energies, and e is the charge on the electron, $C_{\Sigma 1,2}$ are the sum of all capacitances connected to the corresponding island including the coupling capacitance C_m and $n_{g1,2} = (C_{g1,2} V_{g1,2} + C_p V_p) / 2e$ are the normalized charges induced on the corresponding qubit by the d.c. and pulse gate electrodes. The coupling energy E_m depends not only on C_m , but also on the total capacitance of the boxes: $E_m = 4e^2 C_m / (C_{\Sigma 1} C_{\Sigma 2} - C_m^2)$. Application of gate voltages allows us to control diagonal elements of the hamiltonian of equation (1). The circuit was fabricated to have the following relation between the characteristic energies: $E_{J1,2} \approx E_m < E_{c1,2}$. This ensures coherent superposition of the four charge states $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$ around $n_{g1} = n_{g2} = 0.5$, while other charge states are separated by large energy gaps. The above condition justifies the use of a four-level approximation for the description of the system. In our notation $|n_1 n_2\rangle$ of the charge states used throughout the text, n_1 and n_2 refer to the number of excess Cooper pairs in the first and the second qubits, respectively.

In the absence of Josephson coupling, the ground-state charging diagram (n_1, n_2) (ref. 14; Fig. 2a) consists of hexagonal cells whose boundaries delimit two neighbouring charge states with degenerate electrostatic energies. For example, points R and L in Fig. 2a

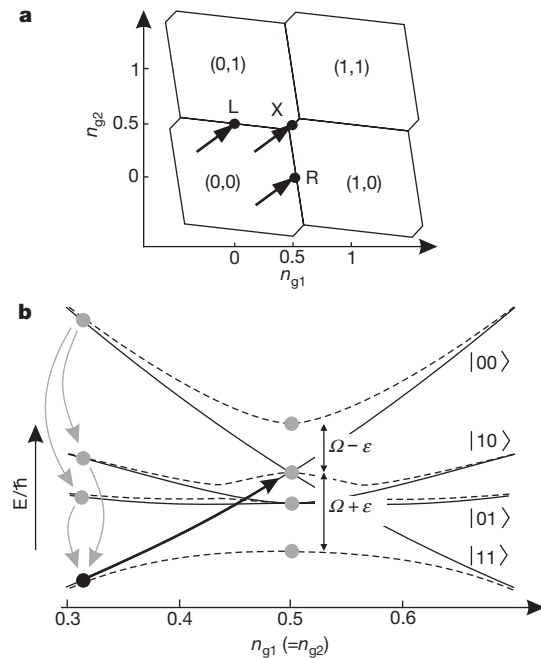


Figure 2 Pulse operation of device. **a**, Ground-state charging diagram of the coupled qubits as a function of the normalized gate charges n_{g1} and n_{g2} . The number of Cooper pairs n_1 and n_2 in the neighbouring cells differs by one. The electrostatic energies $E_{n_1 n_2}$ are degenerate at the boundaries. Points R and L correspond to energy degeneracy in the first and the second qubit, respectively. Point X is doubly degenerate: $E_{00} = E_{11}$ and $E_{10} = E_{01}$. Arrows show how pulses shift the system in the experiment. **b**, Energy diagram of the system along the line $n_{g1} = n_{g2}$. Solid lines are the electrostatic energies of the charge states $|00\rangle$, $|10\rangle$, $|01\rangle$ and $|11\rangle$. Dashed lines are eigenenergies of the hamiltonian equation (1). Far from co-resonance (point X in **a**), the system stays in $|00\rangle$. After the pulse brings the system to the co-resonance (solid arrow), the system starts to evolve producing a superposed state $|\psi(t)\rangle = c_1|00\rangle + c_2|10\rangle + c_3|01\rangle + c_4|11\rangle$. The amplitudes $|c_i|$ ($i = 1, 2, 3, 4$) remain 'frozen' after the pulse termination until the resulting state decays into the ground state. The decay process indicated by grey arrows contributes to the probe currents proportional to the probabilities, equation (3).

correspond to a degeneracy between the states $|00\rangle$ and $|10\rangle$ and the states $|00\rangle$ and $|01\rangle$ differing by one Cooper pair in the first and the second Cooper-pair box, respectively. If we choose the d.c. gate charges n_{g1} and n_{g2} far from the boundaries but within the (0,0) cell, then because of large electrostatic energies we can assume that the system remains in the state $|00\rangle$. As the pulse gate has equal coupling to each qubit, the application of a pulse shifts the state of the system on this diagram along a line tilted at 45° , indicated by arrows in Fig. 2a. The charging diagram remains valid for small Josephson coupling except on the boundaries where charge states become superposed. When the system is driven non-adiabatically to point R or L, it behaves like a single qubit oscillating between the degenerate states with a frequency $\omega_{1,2} = E_{J1,2}/\hbar$. Applying arrays of pulses and measuring oscillations of the probe currents I_1 and I_2 , we can determine the Josephson energies of each qubit. The accuracy of the measured $E_{J1,2}$ is very high, because the electrostatic coupling through C_m has almost no effect on $\omega_{1,2}$ along the boundaries in the vicinity of R and L.

At the 'co-resonance' point $X(n_{g1} = n_{g2} = 0.5)$, the system has a double degeneracy, $E_{00} = E_{11}$, $E_{10} = E_{01}$, and the dynamics of the quantum evolution become more complex and reflect the coupling between the qubits. The cross-section of the energy bands through point X is shown in Fig. 2b. Exactly at the co-resonance, all four

charge states are mixed and the state of the system can be expressed in general as

$$|\psi(t)\rangle = c_1|00\rangle + c_2|10\rangle + c_3|01\rangle + c_4|11\rangle \quad (2)$$

where $|c_i|(i = 1, 2, 3, 4)$ are the time-dependent probability amplitudes obeying a normalization condition $\sum_{i=1}^4 |c_i|^2 = 1$. Using the hamiltonian of equation (1) and initial conditions, we can calculate the probabilities $|c_i|^2$ of each charge state. However, in our read-out scheme, we measure a probe current $I_{1,2}$ proportional to the probability $p_{1,2}(1)$ of each qubit having a Cooper pair on it regardless of the state of the other qubit; that is, $I_1 \propto p_1(1) \equiv |c_2|^2 + |c_4|^2$ and $I_2 \propto p_2(1) \equiv |c_3|^2 + |c_4|^2$. Assuming that the initial state at $t = 0$ is $|00\rangle$, we can derive the time evolution of these probabilities for an ideal rectangular pulse shape of length Δt :

$$p_{1,2}(1) = (1/4)[2 - (1 - \chi_{1,2})\cos\{(\Omega + \varepsilon)\Delta t\} - (1 + \chi_{1,2})\cos\{(\Omega - \varepsilon)\Delta t\}] \quad (3)$$

where $\chi_{1,2} = (E_{J2,1}^2 - E_{J1,2}^2 + E_m^2/4)/(4\hbar^2\Omega\varepsilon)$, $\Omega = ((E_{J1} + E_{J2})^2 + (E_m/2)^2)^{1/2}/2\hbar$, $\varepsilon = ((E_{J1} - E_{J2})^2 + (E_m/2)^2)^{1/2}/2\hbar$.

Unlike the single qubit case, there are two frequencies present in the oscillation spectrum of the qubits: $\Omega + \varepsilon$ and $\Omega - \varepsilon$, both dependent on E_{J1} , E_{J2} and E_m . We can identify these frequencies with the energy gaps in Fig. 2b. Note that in the uncoupled situation ($E_m = 0$), $\Omega \pm \varepsilon = E_{J1,2}/\hbar$ and each qubit oscillates with its own frequency $\omega_{1,2}$. We stress, however, that the above consideration is valid only in the ideal case when the pulse has zero rise/fall time, and the time evolution occurs exactly at the co-resonance point.

The idea of our experiment is shown schematically in Fig. 2b. From the state $|00\rangle$ (shown as a black dot), the pulse (solid arrow) brings the system to the co-resonance point, and the system evolves for the pulse duration time Δt , producing a superposed state equation (2) indicated by grey circles. After the pulse termination, the system remains in the superposed state until it decays (grey arrows) in the ground state by emitting quasiparticles into the probe junctions biased at $V_{b1,2} \approx 600 \mu\text{eV}$. To accumulate a signal, a pulse array ($\sim 3 \times 10^5$ pulses) was applied to the pulse gate. The repetition time between the pulses was 64 ns, long enough (in comparison to the quasiparticle relaxation time of ~ 10 ns) to let the system decay through a Josephson-quasiparticle cycle¹⁵ and give rise to a probe current proportional to $p_{1,2}$. The estimated amplitude of the applied pulses is $V_p \approx 30$ mV.

The results obtained in this way are presented in Fig. 3. First, by tuning n_{g1} and n_{g2} , we do single qubit measurements by bringing

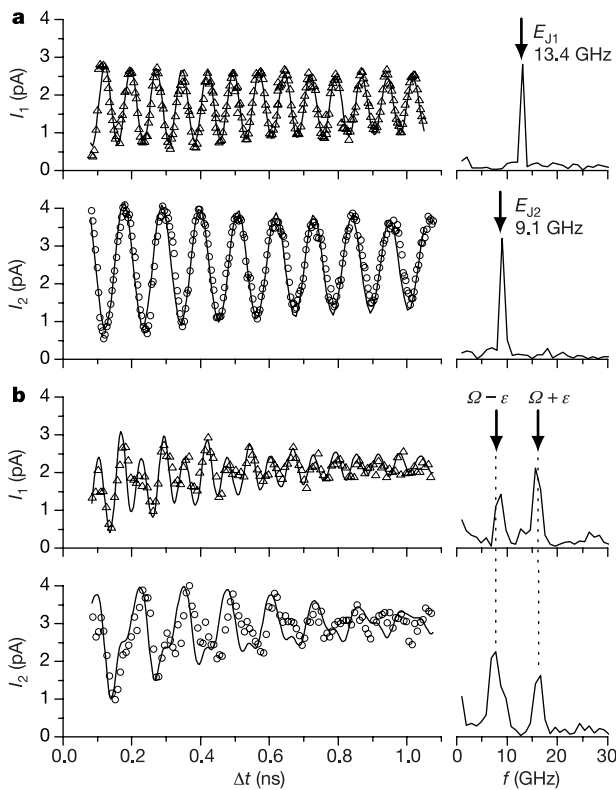


Figure 3 Quantum oscillations in qubits. **a**, Probe current oscillations in the first (top) and the second (bottom) qubit when the system is driven to the points R and L, respectively. Right panel shows corresponding spectra obtained by Fourier transform. In both cases, the experimental data (open triangles and open dots) can be fitted to a cosine dependence (solid lines) with an exponential decay with 2.5 ns time constant. **b**, Probe current oscillations in the qubits at the co-resonance point X. Their spectra (right panel) contain two components. Arrows and dotted lines indicate the position of $\Omega + \varepsilon$, $\Omega - \varepsilon$ obtained from equation (3) using $E_{J1} = 13.4$ GHz, $E_{J2} = 9.1$ GHz measured in the single qubit experiments (Fig. 3a) and $E_m = 15.7$ GHz estimated independently from d.c. measurements. Solid lines are fits obtained from numerical simulation with the parameters $E_{J1} = 13.4$ GHz, $E_{J2} = 9.1$ GHz and $E_m = 14.5$ GHz. Finite pulse rise/fall time and an initial condition that is not pure $|00\rangle$ were taken into account. The introduced exponential decay time is 0.6 ns.

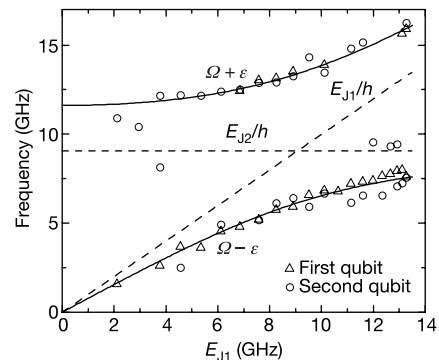


Figure 4 E_{J1} dependence of the spectrum components obtained by Fourier transform of the oscillations at the co-resonance. Open triangles and open circles show respectively frequency components measured in the first and second qubit; solid lines, dependence of $\Omega + \varepsilon$ and $\Omega - \varepsilon$ obtained from equation (3) using $E_{J2} = 9.1$ GHz and $E_m = 14.5$ GHz and varying E_{J1} from zero up to its maximum value of 13.4 GHz; dashed lines, dependence of the oscillation frequencies of both qubits in the case of zero coupling ($E_m = 0$).

the system to point R or L and thus exciting autonomous oscillations in one of the qubits (Fig. 3a). The spectra of the oscillations can be fitted to a cosine function with an exponential decay time of about 2.5 ns. The spectra of the oscillations (right panels of Fig. 3a) obtained by Fourier transform contain one pronounced component at 13.4 GHz for the first qubit and at 9.1 GHz for the second qubit. We identify these values with E_{J1} and E_{J2} . Judging from our previous experiments (see, for example, ref. 13) we conclude that these values are close to what we expect for the given fabrication parameters (that is, overlap area and oxidation conditions). Then, by changing n_{g1} and n_{g2} , the system is driven to the co-resonance point, and the induced quantum oscillations are traced using the same technique. The oscillation pattern becomes more complex (Fig. 3b), and more frequency components appear in the spectrum. The observed spectral properties of the oscillations agree with the predictions of equation (3), in that there are two peaks in the spectrum and the peak positions are close to the expected frequencies $\Omega + \varepsilon$ and $\Omega - \varepsilon$ for the parameters $E_{J1} = 13.4$ GHz and $E_{J2} = 9.1$ GHz measured in the single qubit experiments (Fig. 3a), and $E_m = 15.7$ GHz estimated from independent measurements of d.c. current–voltage–gate–voltage characteristics. The positions of the $\Omega + \varepsilon$ and $\Omega - \varepsilon$ peaks expected from equation (3) are indicated by arrows and dotted lines in Fig. 3. The decay time (~ 0.6 ns) of the coupled oscillations is shorter compared to the case of the independent oscillations, as is expected because an extra decoherence channel appears for each qubit after coupling it to its neighbour. However, the amplitudes of the spectral peaks do not exactly agree with equation (3). We attribute this to the non-ideal pulse shape (finite rise/fall time ~ 35 ps), and the fact that a small shift of n_{g1} and n_{g2} off the co-resonance drastically changes the oscillation pattern. Also, even far from the co-resonance, we still have a small contribution to the initial state from charge states other than $|00\rangle$ distorting the oscillations. We have performed numerical simulations of the oscillation pattern, taking into account a realistic pulse shape and an initial condition of not pure $|00\rangle$, assuming the system is exactly at the co-resonance. The resulting fits are shown in Fig. 3b as solid lines. We found that $E_m = 14.5$ GHz, close to the value estimated from the d.c. measurements, gives better agreement of the fit with the experimental data.

Finally, we checked the dependence of the oscillation frequencies on E_{J1} , controlled by a weak magnetic field (up to 20 G). The results are shown in Fig. 4. The plot contains the data from both qubits represented by open triangles (first qubit) and open circles (second qubit). Without coupling ($E_m = 0$), the single peaks in each qubit would follow the dashed lines with an intersection at $E_{J1} = E_{J2}$. The introduced coupling modifies this dependence by creating a gap, and shifting the frequencies to higher and lower values; the spacing between the two branches is equal to $E_m/2h$ when $E_{J1} = E_{J2}$. We compare the observed dependence with the prediction of equation (3) (given by solid lines) and find a remarkable agreement.

The observed quantum coherent dynamics of coupled qubits in the vicinity of the co-resonance (in particular, the double-frequency structure of the probability oscillations in both qubits, and frequency ‘repulsion’ at $E_{J1} \approx E_{J2}$ —see Figs 3b and 4) indicates that the two qubits become entangled during the course of coupled oscillations, although direct measurement of the degree of entanglement was not possible. Simple calculations based on the standard expression for the entanglement of the pure states¹⁶ show that, with an ideal pulse shape and the $|00\rangle$ initial condition, the wavefunction shown in equation (2) evolves through a maximally entangled state in the case of equal Josephson energies. The numerical simulations confirm that the amount of entanglement does not decrease significantly when realistic experimental conditions are taken into account. The relatively large observed oscillation amplitude (about 50% of the expected value) also suggests the existence of entangled states even in our multi-pulse averaged experiment. □

Received 21 October; accepted 10 December 2002; doi:10.1038/nature01365.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
- Clark, R. G. (ed.) *Experimental Implementation of Quantum Computation* (Rinton, Princeton, 2001).
- Averin, D. V. Quantum computing and quantum measurement with mesoscopic Josephson junctions. *Fortschr. Phys.* **48**, 1055–1074 (2000).
- Makhlin, Y., Schön, G. & Shnirman, A. Quantum state engineering with Josephson-junction devices. *Rev. Mod. Phys.* **73**, 357–400 (2001).
- Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
- Vion, D. et al. Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
- Yu, Y., Han, S., Chu, X., Chu, S.-I. & Wang, Z. Coherent temporal oscillations of macroscopic quantum states in a Josephson junction. *Science* **296**, 889–892 (2002).
- Martinis, J. M., Nam, S., Aumentado, J. & Urbina, C. Rabi oscillations in a large Josephson-junction qubit. *Phys. Rev. Lett.* **89**, 117901 (2002).
- Friedman, J. R., Patel, V., Chen, W., Toltygo, S. K. & Lukens, J. E. Quantum superposition of distinct macroscopic states. *Nature* **406**, 43–46 (2000).
- Van der Wal, C. H. et al. Quantum superposition of macroscopic persistent-current states. *Science* **290**, 773–777 (2000).
- Bouchiat, V., Vion, D., Joyez, P., Esteve, D. & Devoret, M. H. Quantum coherence with a single Cooper pair. *Phys. Scripta T* **76**, 165–170 (1998).
- Flees, D. J., Han, S. & Lukens, J. E. Interband transitions and band gap measurements in Bloch transistors. *Phys. Rev. Lett.* **78**, 4817–4820 (1997).
- Nakamura, Y., Chen, C. D. & Tsai, J. S. Spectroscopy of energy-level splitting between two macroscopic quantum states of charge coherently superposed by Josephson coupling. *Phys. Rev. Lett.* **79**, 2328–2331 (1997).
- Esteve, D. in *Single Charge Tunneling* (eds Grabert, H. & Devoret, M. H.) 109–137 (Plenum, New York, 1992).
- Fulton, T. A., Gammel, P. L., Bishop, D. J., Dunkleberger, L. N. & Dolan, G. J. Observation of combined Josephson and charging effects in small tunnel junction circuits. *Phys. Rev. Lett.* **63**, 1307–1310 (1989).
- Bennett, C. H., Bernstein, H. J., Popescu, S. & Schumacher, B. Concentrating partial entanglement by local operations. *Phys. Rev. A* **53**, 2046–2052 (1996).

Acknowledgements We thank B. L. Altshuler, X. D. Hu, H. Im, S. Ishizaka, F. Nori, T. Sakamoto and J. Q. You for discussions. D.V.A. was supported by AFORS, and by NSA and ARDA under an ARO contract.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.T. (e-mail: tsai@frl.cl.nec.co.jp).

Probing molecular dynamics with attosecond resolution using correlated wave packet pairs

Hiromichi Niikura*, F. Légaré†, R. Hasbani*, Misha Yu Ivanov*, D. M. Villeneuve* & P. B. Corkum*

* National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada

† Université de Sherbrooke, Sherbrooke PQ, Canada

Spectroscopic measurements with increasingly higher time resolution are generally thought to require increasingly shorter laser pulses, as illustrated by the recent monitoring of the decay of core-excited krypton¹ using attosecond photon pulses^{2,3}. However, an alternative approach to probing ultrafast dynamic processes might be provided by entanglement, which has improved the precision^{4,5} of quantum optical measurements. Here we use this approach to observe the motion of a D_2^+ vibrational wave packet formed during the multiphoton ionization of D_2 over several femtoseconds with a precision of about 200 attoseconds and 0.05 Ångströms, by exploiting the correlation between the electronic and nuclear wave packets formed during the ionization event. An intense infrared laser field drives the electron wave packet, and electron recollision^{6–11} probes the nuclear motion. Our results show that laser pulse duration need

not limit the time resolution of a spectroscopic measurement, provided the process studied involves the formation of correlated wave packets, one of which can be controlled; spatial resolution is likewise not limited to the focal spot size or laser wavelength.

Our experiment is analogous to conventional pump-probe measurements¹², but the pump and probe occur within one optical cycle, a process that we call 'sub-laser-cycle molecular dynamics' (Fig. 1a). Ionization, which forms correlated wave packets around each peak of the laser field, is the pump. By removing one electron (creating an electron wave packet in the continuum), we weaken the force binding the protons and therefore launch a correlated vibrational wave packet (Fig. 1b). Because of its small mass, only the electron wave packet is influenced by the laser field. In linear polarization, the electron wave packet is first moved away from the parent ion but is pulled back by the laser field. The probability of electron re-collision with the parent ion reaches a maximum at a well-defined laser phase, about two-thirds of an optical period after the electron's transition to the continuum (Fig. 1a). Re-collision probes the vibrational wave packet (Fig. 1b). Changing the laser wavelength delays re-collision just as changing the position of a translation stage changes an optical pump-probe delay. (Neither re-collision nor the pump-probe analogy is essential: fast measurements are possible in their absence if both correlated partners are controlled, as discussed below.)

Re-collision between an electron and its parent ion has long been known as a source of high-harmonics emission^{7,8,13,14} and generation of high-energy electrons^{7,9}. It is also responsible for correlated multi-electron ionization in strong laser fields^{7,10,13}, fragmentation in small molecules¹¹ and attosecond pulse generation^{2,3,15}. For molecules, re-collision will also imprint the spatial structure of the ion on the harmonic emission spectrum¹⁶ or on the photo-

electron spectrum⁶ in analogy with conventional electron diffraction (see, for example, ref. 17).

We use 40-fs laser pulses with an intensity of $1.5 \times 10^{14} \text{ W cm}^{-2}$ tunable from 800 to 1,850 nm. Ionization induces vertical population transfer from the D_2 ($X^1\Sigma_g^+$) potential energy surface to the D_2^+ ($X^2\Sigma_g^+$) surface, with less than 10^{-6} of the population transferred to any other levels of D_2^+ (I. Litvinyuk, P. Dooley, D.M.V. and P.B.C., manuscript in preparation). Because the equilibrium internuclear separation is greater for D_2^+ (1.06 Å) than for D_2 (0.74 Å), ionization initiates a vibrational wave packet motion (Fig. 1b). For all wavelengths that we use, the instantaneous ionization rate¹⁸ is sharply peaked near the instantaneous maxima of the oscillating field. The ionization bursts last 190–300 as (800 nm to 1.85 µm). The nuclei are essentially frozen by their inertia on that timescale, ensuring that the pump stage is the same for all wavelengths.

We use electron-ion re-collision to observe motion on the D_2^+ ($X^2\Sigma_g^+$) surface. Although the electron returns to the parent ion

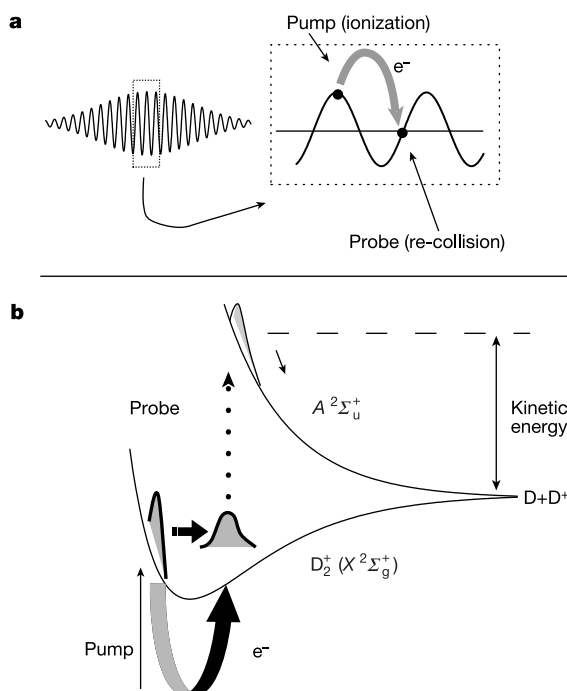


Figure 1 Processes probed and exploited with the sub-laser-cycle dynamics method using correlated wave packet pairs. **a**, Ionization forms correlated electronic and vibrational wave packets near each peak of the laser field. **b**, The vibrational wave packet moves on the D_2^+ ($X^2\Sigma_g^+$) surface while the laser field causes the electron wave packet to re-collide (represented by a thick arrow in **a** and **b**). Inelastic scattering during re-collision probes the vibrational wave packet's position. We measure the kinetic energy of the D^+ fragments (**b**).

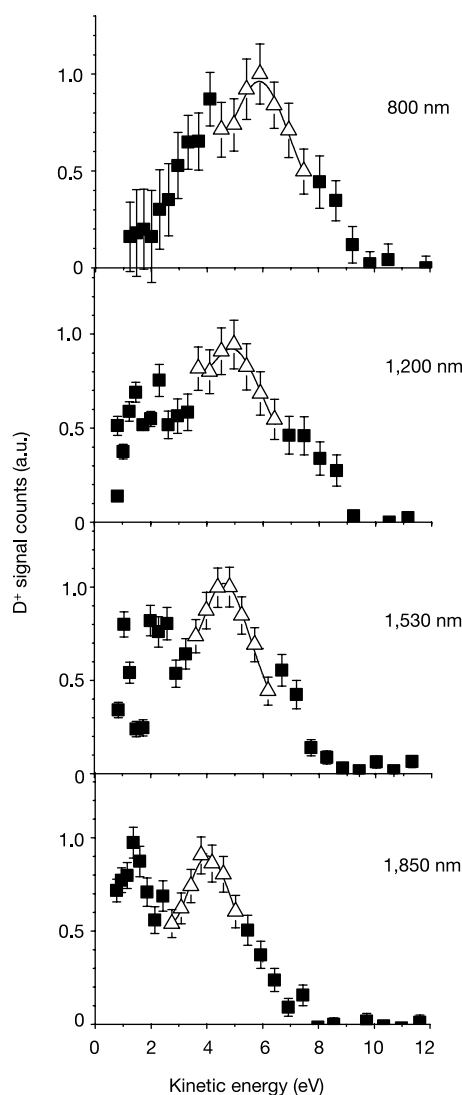


Figure 2 Kinetic energy distribution of D^+ at different pump-probe delay times. Shown is the experimental kinetic energy distribution of D^+ obtained using four different laser wavelengths, corresponding to different pump-probe delay times. Experimental error is estimated for each point by taking the square root of the signal counts. We identify the largest high-energy peak by the open data points with dissociation from D_2^+ ($A^2\Sigma_u^+$). The curve is obtained by a fit to the data as described in the text.

several times after ionization, the first return of the electron wave packet dominates⁶ and can thus be used as a probe pulse with a duration of about 1 fs at 800 nm (ref. 6). The time delay between the creation of the correlated wave packets and their re-collision (pump-probe delay) is controlled by varying the laser wavelength. Our delay times range from 1.7 to 4.2 fs (about two-thirds of the period for $\lambda = 800\text{--}1,850\text{ nm}$).

Inelastic scattering of the electron with the parent ion leads to excitation or double ionization, giving rise to the dissociative fragments of D^+ . The kinetic energy of the fragments is determined by the internuclear separation at the time of electron re-collision. By using a small aperture in the time-of-flight apparatus, we observe only those fragments that originate from molecules aligned perpendicular to the laser polarization. This configuration eliminates laser-induced coupling between $X^2\Sigma_g^+$ and $A^2\Sigma_u^+$. Our choice of alignment ensures that the wave packet motion is simple, that it can be easily modelled and that $A^2\Sigma_u^+$ is a good reference state. The existence of the well-understood reference state allows us to unequivocally interpret the kinetic energy spectrum and use it to measure the position of the vibrational wave packet.

Figure 2 shows the kinetic energy spectrum of D^+ that is due to re-collision, for all wavelengths that we have studied. Although many dissociation pathways contribute to the observed kinetic energy spectra, we can isolate the Σ_u state as particularly significant. This is because, among all inelastic scattering events from the Σ_g surface, excitation to Σ_u has the highest cross-section over the energy range examined here¹⁹. In addition, the Σ_u leads to fragments with the second-largest kinetic energy. Therefore this channel is easily identified in Fig. 2. We have labelled the data points that we associate with this channel with triangles. In Fig. 2 we see the motion of the wave packet reflected in the shift of this peak to lower energy with longer wavelength light.

To convert the measured kinetic energy spectrum to the vibrational wave packet on the Σ_g surface, we use the reflection principle²⁰. The transformation includes the radial dependence of the collision cross-section, and a self-consistent addition of the kinetic energy of the $D_2^+ \Sigma_g^+$ wave packet, assuming that the initial average wave packet velocity is zero and initial average kinetic energy is given by the zero-point energy of D_2 . We then fit the wave packet with a gaussian. Its projection back to the kinetic energy spectrum yields the solid curves in Fig. 2. The gaussian fit on Σ_g

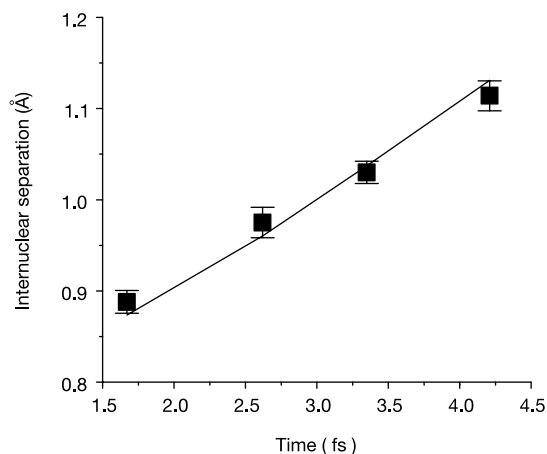


Figure 3 The measured wave packet position as a function of time. To associate the laser wavelength with the time of re-collision, we use the mean time of re-collision for the first electron micro-bunch at each laser wavelength. The position is determined from the peak of the fit in Fig. 2. The experimental error bars are determined by the error of the fit. The solid line is the calculated nuclear position as a function of time for a wave packet evolving on the $D_2^+ (X^2\Sigma_g^+)$ state. Because the time window that we measure is only about one-quarter of the full vibrational period of $D_2^+ (X^2\Sigma_g^+)$, the motion is almost linear.

determines the experimentally measured wave packet and the position of its peak. We emphasize that this is a fit to the experimental data, not a theoretical curve.

Figure 3 is a plot of the experimentally determined internuclear separation as a function of time. We represent the position of the vibrational wave packet by the peak position of the gaussian at each laser wavelength. At each wavelength, the re-collision time is defined by the mean time for the electron wave packet's first return. This establishes the experimental points in Fig. 3. The error bars are determined by the deviation of the fit. We determine the position and the velocity of the wave packet with sub-fs, sub-Å resolution.

Because we are demonstrating a new technique for ultrafast measurement, it is important to compare our measurement against a standard. The solid line in Fig. 3 is the result of a simulation of the motion of a vibrational wave packet on the Σ_g surface by solving the time-dependent Schrödinger equation. The discrepancy is within the error bars.

The portion of the kinetic energy distribution that is higher than that assigned to the Σ_u state in Fig. 2 corresponds to transitions to the D_2^{++} state. For 800 nm the re-collision kinetic energy is insufficient to ionize D_2^+ , but at other wavelengths a contribution can be seen. The peak at lower kinetic energy is not clearly identified. It contains contributions of other excited states and of the electron's multiple returns⁶. These are also seen to move to lower kinetic energies as the re-collision time is delayed.

Our current measurement accuracy is limited by noise, but noise in our data is not a fundamental limit. Using three-dimensional imaging detectors instead of observing through a small pinhole would increase our signal strength by about two orders of magnitude. A high-density jet source and a laser with higher repetition rate could further increase our signal by orders of magnitude.

Our method is based on the principle of creating two correlated wave packets that could be formed on any crest of the laser field. We assert that these wave packets are actually entangled. A field crest occurs once during each half-period throughout the 40-fs pulse. Unless the electron and vibrational wave packet are entangled, the kinetic energy spectrum for D^+ must have a 3-eV modulation (for 800 nm light, 1.5 eV per deuteron). Such modulation is well known in strong-field single-particle processes such as high harmonics generated by ionizing atomic media^{8,9,13,14}.

In fact, there is no 1.5 eV modulation in the kinetic energy spectrum at 800 nm in Fig. 2. We measured the spectrum with higher resolution to seek such modulation; none was observed. These experimental results are consistent with the vibrational and electron wave packets being entangled. Entanglement arises because the departing electron contains information on the nuclear coordinate at the time of ionization and on the excited state of the D_2^+ formed in the re-collision. By observing only the deuterium ions, we integrate over all final states of other unseen, but entangled, partner particles. Such partial measurement eliminates the interference.

Regarding future applications, we note that the ultimate time resolution of correlated measurements is determined by the energy bandwidth of the electron bunch—many tens of electron volts in the present experiment. To obtain a resolution of 200 as will be a challenge similar to that of producing the shortest optical pulses².

Also, as long as both particles can be controlled, re-collision is unnecessary. For example, if two electrons are ejected into a field in a correlated fashion, the time difference between their ejection is mapped onto the final energy and angular spectra^{10,15,21}.

Correlated particles produced in a process with important dynamics to probe can always be measured as long as at least one of the pair can be controlled. Because of their low mass, electrons are easily controlled, and at progressively higher intensities, muons, protons, alpha particles and ions are also controllable. Applying correlation to nuclear dynamics requires a means of stimulating a process such that it will occur with high probability during the duration of the intense ultrashort pulse initiating it. Nuclear

processes can be stimulated in an analogous manner to inelastic scattering in our experiment—by using a precursor molecule (such as HCl) and forcing ‘re-collision’ between a proton and the heavy nucleus with an intense few-cycle optical pulse. To follow the subsequent dynamics, the relative trajectories of the correlated, charged nuclear fragments, as influenced by the field, can be measured.

Regarding the ‘probe’, we note that any method of observing the relative evolution of the correlated particles can be used for measurement. In molecular science, harmonic generation is one possible method¹⁶ where phase matching eliminates the contribution from all but the first electron wave packet return, which occurs about two-thirds of a period following ionization. Elastic scattering is a second possible method⁶, with diffraction determining the nuclear position at the time of re-collision. □

Methods

Production and control of the electron wave packet

An optical parametric amplifier is used to shift the 40-fs output of a Ti:sapphire laser system to longer wavelengths. We use 800 nm, 1.2 μm , 1.53 μm and 1.85 μm pulses. Each ionizes D₂ near the field maximum and provides a progressively longer time delay between ionization and re-collision. Taking the peak of the first electron microbunch to be the delay time, this corresponds to delay times of 1.7, 2.7, 3.4 and 4.2 fs. Changing the wavelength has other implications. The intensity and wavelength determine the maximum re-collision energy $3.17q^2E^2/(4m\omega^2)$ of the electron⁷ (q and m are the electron mass and charge, ω and E are the laser pulse’s angular frequency and electric field amplitude at the time of ionization). We use a light intensity of $1.5 \times 10^{14} \text{ W cm}^{-2}$, calibrated against the ionization of xenon. Whereas at 800 nm the peak kinetic energy of the re-colliding electron is 30 eV, the ω^{-2} scaling of the kinetic energy means that at 1.85 μm the energy of the peak re-colliding electron is about 160 eV.

Kinetic energy analysis

The kinetic energy spectrum of D⁺ was measured with a time-of-flight (TOF) mass spectrometer filled with 10^{-6} torr of D₂. The TOF axis was perpendicular to the direction of propagation of the laser beam. A 1-mm-diameter hole in the electrode placed 1.5 cm from the laser focus selects only those D⁺ ions resulting from dissociation of D₂⁺ molecules that are aligned along the TOF axis. For the high-resolution results taken with 800-nm light (not shown), the extraction field was 133 V cm⁻¹ and the acceptance angle of the TOF was 6 degrees at 8 eV. For the lower-resolution results, the extraction field was 400 V cm⁻¹ giving an acceptance angle of 9 degrees at 8 eV.

Selection of the re-collision channel

Deuterium ions are produced by a number of processes during strong field ionization of D₂. We distinguish fragments resulting from inelastic scattering caused by the returning electron (Fig. 2) from those produced by any other process (such as sequential double ionization²², bond softening²³ and enhanced ionization^{24–27}) using the strong sensitivity of recollision phenomena to the ellipticity of the light polarization^{6,7,13}. It only takes a small ellipticity to displace the electron laterally with respect to the parent ion so recollision is impossible⁷. In strong fields, the difference between the spectrum obtained with linear and elliptically polarized light is due to recollision.

Received 24 July 2002; accepted 13 January 2003; doi:10.1038/nature01430.

- Drescher, M. *et al.* Time-resolved atomic inner-shell spectroscopy. *Nature* **419**, 803–807 (2002).
- Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
- Paul, P. M. *et al.* Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–1692 (2001).
- D’Ariano, G. M., Presti, P. L. & Paris, M. G. A. Using entanglement improves the precision of quantum measurements. *Phys. Rev. Lett.* **87**, 27040 (2001).
- Resch, K. J., Lundeen, J. S. & Steinberg, A. M. Experimental observation of nonclassical effects on single-photon detection rates. *Phys. Rev. A* **63**, 020102 (2000).
- Niikura, H. *et al.* Sub-laser-cycle electron pulses for probing molecular dynamics. *Nature* **417**, 917–922 (2002).
- Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Dietrich, P., Burnett, N. H., Ivanov, M. & Corkum, P. B. High harmonic generation and correlated two-electron multiphoton ionization with elliptically polarized light. *Phys. Rev. A* **50**, 3585–3588 (1994).
- Paulus, G. G., Nicklich, W., Xu, H., Lambropoulos, P. & Walther, H. Plateau in above threshold ionization spectra. *Phys. Rev. Lett.* **72**, 2851–2854 (1994).
- Weber, Th. *et al.* Correlated electron emission with multiphoton double ionization. *Nature* **405**, 658–661 (2000).
- Bhardwaj, V. R., Rayner, D. M., Villeneuve, D. M. & Corkum, P. B. Quantum interference in double ionization and fragmentation of C₆H₆ in intense laser fields. *Phys. Rev. Lett.* **87**, 253003 (2001).
- Zewail, A. H. Femtosecond chemistry. *Science* **242**, 1645–1653 (1988).
- Krause, J. L., Schafer, K. J. & Kulander, K. C. High-order harmonic generation from atoms and ions in the high intensity regime. *Phys. Rev. Lett.* **68**, 3535–3538 (1992).
- Lewenstein, M., Balcou, Ph., Ivanov, M. Yu., L’Huillier, A. & Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).
- Ivanov, M., Corkum, P. B., Zuo, T. & Bandrauk, A. Routes to control of intense-field atomic polarizability. *Phys. Rev. Lett.* **74**, 2933–2936 (1995).

- Lein, M., Hay, N., Velotta, R., Marangos, J. P. & Knight, P. L. Role of the intramolecular phase in high-harmonic generation. *Phys. Rev. Lett.* **88**, 183903 (2002).
- Hargittai, I. & Hargittai, M. *Stereochemical Applications of Gas-Phase Electron Diffraction* (VCH, New York, 1998).
- Yudin, G. L. & Ivanov, M. Yu. Physics of correlated double ionization of atoms in intense laser fields: Quasistatic tunneling limit. *Phys. Rev. A* **63**, 033404 (2001); erratum **64**, 019902 (2001).
- Peek, J. M. Inelastic scattering of electrons by the hydrogen molecule ion. *Phys. Rev. A* **134**, 877–883 (1964).
- Schinke, R. *Photodissociation Dynamics* 114–115 (Cambridge Univ. Press, Cambridge, UK, 1993).
- Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* **88**, 173903 (2002).
- Lambropoulos, P. Mechanisms for multiple ionization of atoms by strong pulsed lasers. *Phys. Rev. Lett.* **55**, 2141–2144 (1985).
- Zavriyev, A., Bucksbaum, P. H., Muller, H. G. & Schumacher, D. W. Ionization and dissociation of H₂ in intense laser fields at 1.064 μm , 532 nm, and 355 nm. *Phys. Rev. A* **42**, R5500–R5513 (1990).
- Codling, K. & Frasinski, L. J. Dissociative ionization of small molecules in intense laser fields. *J. Phys. B* **26**, 783–809 (1993).
- Seideman, T., Ivanov, M. Yu. & Corkum, P. B. Role of electron localization in intense-field molecular ionization. *Phys. Rev. Lett.* **75**, 2819–2822 (1995).
- Zuo, T. & Bandrauk, A. D. Charge-resonance-enhanced ionization of diatomic molecular ions by intense lasers. *Phys. Rev. A* **52**, R2511–R2514 (1995).
- Constant, E., Stapelfeldt, H. & Corkum, P. B. Observation of enhanced ionization of molecular ions in intense laser fields. *Phys. Rev. Lett.* **76**, 4140–4143 (1996).

Acknowledgements We acknowledge discussions with A. Stolow, A. Sokolov, A. D. Bandrauk, S. Chelkowski and J. Marangos. The authors appreciate financial support from Canada’s Natural Science and Engineering Research Council, the Canadian Institute for Photonics Innovation, and Le Fonds Québécois de la Recherche sur la Nature et les Technologies.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.B.C. (e-mail: Paul.Corkum@nrc.ca).

Multi-colour organic light-emitting displays by solution processing

C. David Müller^{*†}, Aurélie Falcou[‡], Nina Reckefuss^{*†}, Markus Rojahn[§], Valérie Wiederhirm[§], Paula Rudati^{*†}, Holger Frohne^{*†}, Oskar Nuyken[§], Heinrich Becker[‡] & Klaus Meerholz^{*†}

^{*} Department Chemie, Universität München, Butenandtstr. 11, 81377 München, Germany

[‡] Covion Organic Semiconductors GmbH, Industriepark Hoechst, G865, 65926 Frankfurt, Germany

[§] Lehrstuhl für Makromolekulare Stoffe, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany

Organic light-emitting diodes (OLEDs) show promise for applications as high-quality self-emissive displays for portable devices such as cellular phones and personal organizers^{1–4}. Although monochrome operation is sufficient for some applications, the extension to multi-colour devices—such as RGB (red, green, blue) matrix displays—could greatly enhance their technological impact. Multi-colour OLEDs have been successfully fabricated by vacuum deposition of small electroluminescent molecules, but solution processing of larger molecules (electroluminescent polymers) would result in a cheaper and simpler manufacturing process. However, it has proved difficult to combine the solution processing approach with the high-resolution patterning techniques required to produce a pixelated display. Recent attempts have focused on the modification of standard printing techniques, such as screen printing^{5–7} and ink jetting⁸, but those still have technical drawbacks. Here we report a class of electroluminescent polymers that can be patterned in a way similar to standard photoresist materials—soluble polymers with oxetane

[†] Present address: Institut für Physikalische Chemie, Universität zu Köln, Luxemburgerstr. 116, 50939 Köln, Germany.

sidegroups that can be crosslinked photochemically to produce insoluble polymer networks in desired areas. The resolution of the process is sufficient to fabricate pixelated matrix displays. Consecutive deposition of polymers that are luminescent in each of the three RGB colours yielded a device with efficiencies comparable to state-of-the-art OLEDs and even slightly reduced onset voltages.

Small-molecule OLEDs are commonly fabricated by vacuum deposition, whereas polymer OLEDs are usually built by solution processing (mostly spin-coating). The latter route has the advantage of much easier device manufacture, and efficient single-layer devices with low turn-on voltages can be obtained. By contrast, vacuum deposition allows for the fabrication of multiple-layer devices, which are generally more efficient than their single-layer analogues. Furthermore, by using shadow masks, patterned structures can be obtained. Recent research efforts focus on the adaptation of printing techniques such as screen printing^{5,6} and ink jetting⁸ to polymer OLED technology. Screen printing has been demonstrated as a simple patterning method to obtain macro-structured two-colour signs. However, owing to its accuracy limitations, it can currently not be used for full-colour display fabrication. Ink jet printing on the other hand offers the possibility of achieving high resolution, but has the disadvantage that the wettability of the substrate has to be pre-adjusted (patterned by photoresist technique) in order to place the drops at the precise location (wetting of pixel, but not of bank). Often inhomogeneous films are obtained after drying. In addition, the alignment of drops in consecutive depositions is a rather cumbersome process; for example, the deposition of an electroluminescent (EL) polymer on top of the commonly used

hole-injecting polymer PEDOT (poly(3,4-ethylenedioxy-thiophene)). Other potential large-area patterning routes are laser-induced thermal imaging⁹ and reductive laser-induced bleaching¹⁰. The former method transfers the material from a precursor substrate to the desired locations by local heating/contact printing. In the latter case, a mixture of red- and green-emitting EL polymers embedded in a blue-emitting host was used. Selective bleaching yielded the three basic colours. Both processes are very delicate. In particular, laser-induced bleaching most probably suffers from relatively short lifetimes owing to the presence of large amounts of decomposition products.

In this paper we demonstrate a three-colour (RGB) OLED by solution processing using emitter polymers with photoresist properties, that is, soluble polymers, which can be cured photochemically to yield an insoluble form (polymer networks). In earlier work, we have successfully used the crosslinking of oxetane-functionalized hole-conductors for the fabrication of highly efficient blue OLEDs with three graded hole-injection layers^{11,12}. Unlike other approaches in this direction using reactive side- or main-chain polymers with pendent polymerizable moieties^{13–16} or monomers with several reactive groups^{17–19}, the use of oxetane reactive units is the only one so far that not only allows us to deposit an (in principle) unlimited number of layers¹¹, but simultaneously maintains the desired electrical and optical functionalities. Furthermore, the volume shrinkage is small and thus formation of microcracks upon curing is avoided. Here, we expand this approach to light-emissive polymers of the poly-spiro family²⁰.

The synthesis of the three oxetane-functionalized spirobifluorene-co-fluorene polymers emitting blue, green and red light (**P1–P3**) follows state-of-the-art procedure²⁰ and is depicted in Fig. 1. The monomer feed ratios used in the preparation of the polymers are given in Table 1. Thus, in the case of the blue-emitting polymer a true comparison between the crosslinkable polymer **P1** and the non-crosslinkable reference **P4** is possible, because monomer **3** is structurally very similar to monomer **2**, the only difference being the oxetane units. By contrast, no such comparison with their non-crosslinkable state-of-the-art analogues is possible for the green- and red-emitting crosslinkable polymers²⁰, because monomer **2** must replace some of the emissive moieties.

The polymerizations were achieved through a Suzuki polycondensation of the diboronic ester **1** and two or more aromatic dibromides (**2–7**) with 0.8 mol% of tetrakis-(triphenylphosphine)-palladium(0) as catalyst for 48 h at a temperature of 87 °C in a mixture of toluene (with some ethanol) and water with tri-potassium

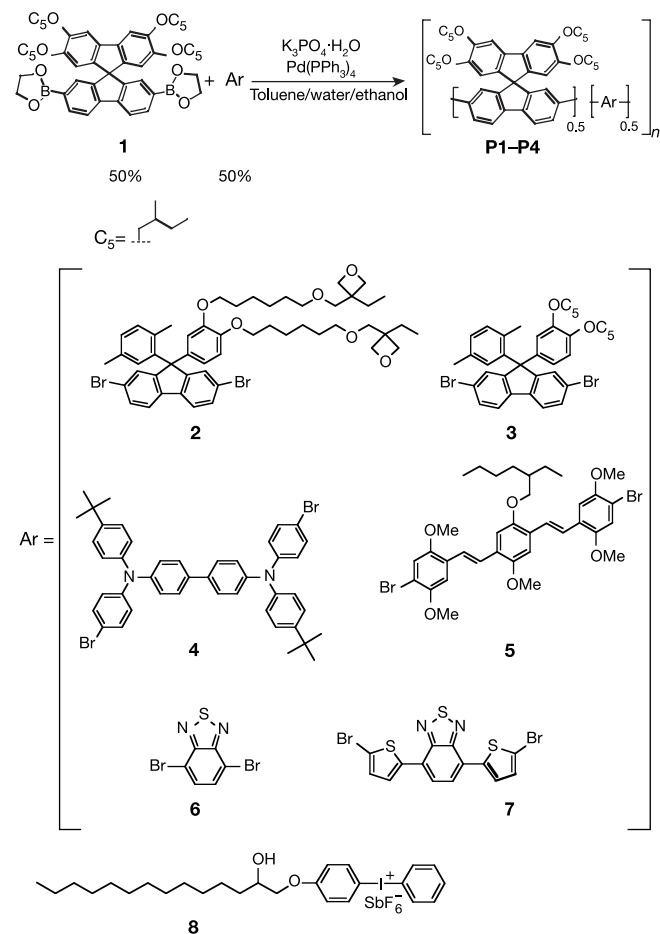


Figure 1 Synthesis of polymers **P1** to **P4**; chemical structure of photoacid **8**.

Table 1 Composition and properties of the studied polymers

	Colour-emitting polymers			
	P1 (X-blue)	P2 (X-green)	P3 (X-red)	P4 (blue)
1 Moiety (mol%)	50	50	50	50
2	25	25	25	–
3	15	–	–	40
4	10	10	10	10
5	–	15	–	–
6	–	–	10	–
7	–	–	5	–
M_w (g mol ^{–1})	115,000	180,000	135,000	195,000
M_n (g mol ^{–1})	48,000	47,000	55,000	72,000
$\lambda_{max,EL}$ (nm)	457	507	650	454
$\lambda_{max,abs}$ (nm)	399	392	394	397
CIE 1931 x	0.16	0.31	0.67	0.15
CIE 1931 y	0.19	0.58	0.33	0.16
η_{max} (pristine) (cd A ^{–1})	2.9	7.0	1.0	3.0
U [V] at 100 cd m ^{–2} (pristine)	4.5	3.8	8.3	4.6
η_{max} (crosslinked) (cd A ^{–1})	3.0	6.5	1.1	n.a.
U [V] at 100 cd m ^{–2} (crosslinked)	4.5	3.8	7.5	n.a.

Monomer feed ratio in the polymerization of the polymers (Fig. 1) and selected physical properties of the investigational polymers. n.a., not applicable. EL, electroluminescence; abs, absorbance; M_n , number average of molecular weight; M_w , weight average of molecular weight; $\lambda_{max,abs}$, maximum wavelength of absorption; $\lambda_{max,EL}$, maximum wavelength of electroluminescence; CIE 1931 x, y, colour coordinates according to the 1931 CIE convention; U [V] at 100 cd m^{–2}, the voltage that needs to be applied in order to achieve 100 cd m^{–2} light output; 'pristine' refers to non-crosslinked samples without addition of the photoacid **8**; X, crosslinkable.

phosphate as the base. It is another advantage of the oxetane functionality that it is chemically stable under these conditions. The polymers were obtained in yields of approximately 75% after precipitation in methanol and subsequent purification. The resulting conjugated polymers contain about 25 mol% of the oxetane-functionalized monomer **2** and are soluble in common organic solvents, such as toluene, tetrahydrofuran (THF) or xylene. The physical properties of these polymers are listed in Table 1.

Oxetanes are commonly polymerized cationically^{21,22}. Therefore, to initiate the crosslinking process the photoacid **8** (Fig. 1) was added. Upon ultraviolet (UV) illumination, the initiator decomposes via a multiple-step mechanism and eventually generates protons H^+ , which open the oxetane ring and start the polymerization²¹. Ultimately, the initially generated protons are consumed in the polymerization. After the softcuring step (see Methods) oxetane groups could no longer be detected by Fourier transform infrared (FTIR) investigations¹¹. The chain end consists of an oxonium cation²³, so the crosslinked films were treated with solutions of various bases and nucleophiles for neutralization purposes. We found that rinsing the films with pure THF led to the best device performance. Further investigations are underway.

The electron-rich, highly fluorescent poly(spirobifluorene-co-fluorenes) are efficient photosensitizers, so in a competitive side

reaction, radical cations are formed in the EL layer due to electron transfer between the polymer and the initiator. These cations effectively quench the EL. However, by a post-baking step at 180 °C, the EL can be completely recovered. The exact mechanism is currently under investigation and will be published elsewhere (C.D.M. and K.M., manuscript in preparation; ref. 21). We point out that with all three polymers, the emission spectra (photoluminescence (PL) as well as EL) were not affected by the crosslinking or the consecutive curing step.

OLED devices were prepared (see Methods for details) with the general structure: ITO/PEDOT (20 nm)/EL polymer (80 nm)/Ca (20 nm)/Ag (200 nm). In the non-crosslinked reference devices no photoinitiator **8** was added to the solution of the EL polymer.

In their non-crosslinked form, all three polymers show a performance that compares favourably with state-of-the-art EL polymers reported²⁰. In particular, the blue-emitting polymer **P1** exhibits an excellent maximum efficiency of $\eta_{\max} \approx 2.9 \text{ cd A}^{-1}$. This compares to $\eta_{\max} \approx 3.0 \text{ cd A}^{-1}$ in the reference devices using non-crosslinkable **P4** (Table 1). The latter fact demonstrates that the incorporation of the oxetane groups did not significantly alter the EL properties of the resulting polymer.

The non-crosslinked polymers **P3** (red; $\eta_{\max} \approx 1 \text{ cd A}^{-1}$) and **P2** (green; $\eta_{\max} \approx 7 \text{ cd A}^{-1}$) exhibit somewhat reduced efficiency compared with previous non-crosslinkable examples (2.0 and 10.0 cd A^{-1} , respectively²⁰). We attribute this (particularly in the red-emitting case) to an unfavourably changed charge-carrier balance in the polymer, resulting from the necessity of incorporating the oxetane-functionalized monomer **2** in 25% feed (partially replacing **6** and **7** in state-of-the-art analogues). Another reason for

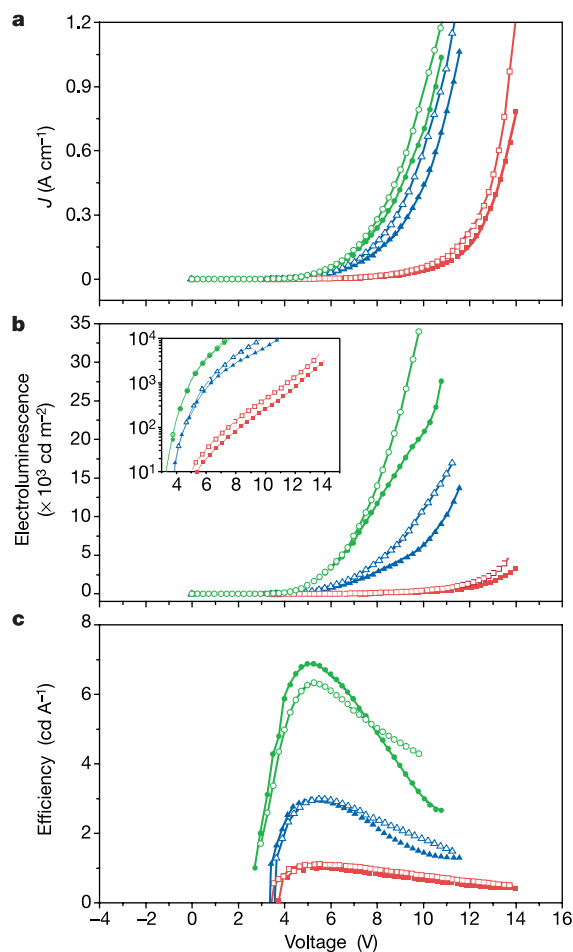


Figure 2 Performance of OLEDs based on crosslinkable polyspiro-derivatives. Voltage dependence of **a**, the current density J , **b**, the electroluminescence intensity, and **c**, the luminous efficiency for red-emitting (**P3**, squares), green-emitting (**P2**, circles), and blue-emitting (**P1**, triangles) devices. The open symbols refer to the crosslinked devices, the solid symbols to the non-crosslinked reference devices. The lines are guides to the eye. The inset of **b** shows the same data sets as the main figure, however, on a semi-logarithmic scale.

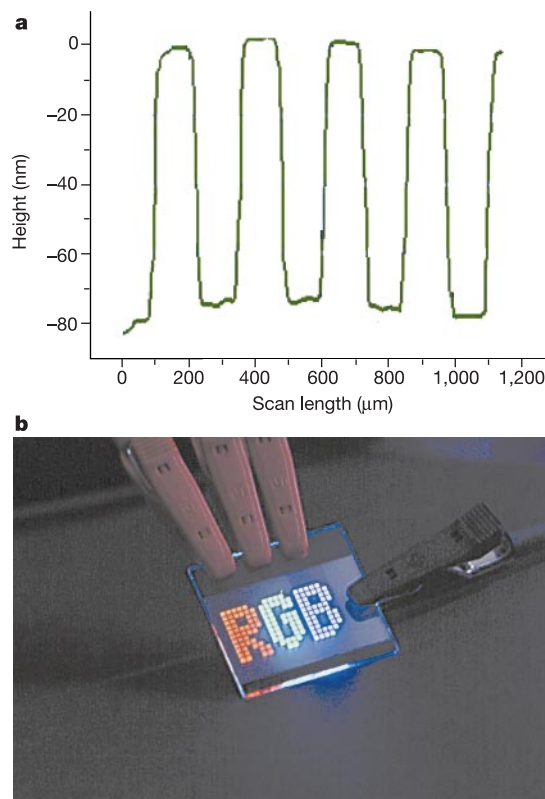


Figure 3 Demonstration of the photoresist properties of crosslinkable polyspiro-derivatives. **a**, Cross-section of a 73-nm-thick photo-structured layer of the blue emitting polymer (**P1**) illuminated through a grid with 250- μm spacing (125- μm openings). **b**, Photograph of a RGB (red, green, blue) device. Dimensions of the glass substrate are 25 $\times$ 25 mm.

the reduced performance is probably the use of calcium cathodes in this work as opposed to barium in ref. 20.

The OLEDs made from crosslinked **P1** and **P3** exhibited identical, if not even slightly improved maximum efficiency $\eta_{\max}$ compared to the non-crosslinked references (Fig. 2; see also Table 1). In contrast, OLEDs made with the crosslinked green-emitting **P2** achieved only about 92% of the non-crosslinked analogues. The turn-on voltage for light emission was slightly reduced in all crosslinked devices. Furthermore, in all three cases the crosslinked devices exhibited higher efficiency at high luminance, and they could be driven to higher current density before breakdown occurred.

After having demonstrated the ability to crosslink the materials without significant alteration of their EL properties, we then took advantage of the photoresist-type properties of the materials and investigated the possibility of photo-patterning the films. For this purpose, the layers were irradiated through a shadow mask (grating; 250- μm period, 125- μm gap size). After curing, the films were developed by immersing them into THF, which re-dissolved the non-irradiated areas of the film. The line scan obtained with a profilometer (Fig. 3a) clearly demonstrated the formation of 125- μm -wide ridges with a height identical to the previous film thickness (that is, full modulation). The edges are sharp, demonstrating that no significant diffusion of the photogenerated protons into the dark regions takes place. We point out that the achievable resolution in our materials is much higher (a few micrometres), as was tested by illuminating with a holographic interference pattern (E. Mecher, C.D.M. and K.M., manuscript in preparation). By applying the above-mentioned procedure three times, once for each polymer, we were able to fabricate pixelated devices with three individually addressable colours (Fig. 3b).

In conclusion, we have demonstrated a RGB polymer OLED obtained by simple spin-coating. Except for the blue-emitting material the efficiencies of the new generation of EL polymers do not reach the best state-of-the-art values quite yet, but this can certainly be achieved by optimizing the synthetic route. Coatings were obtained that can be processed similarly to ordinary photoresists using standard lithographic techniques. We believe this to be a highly promising technology for the fabrication of true-colour matrix displays with many advantages, such as thermal and/or mechanical robustness (due to the formation of polymer networks) and higher efficiency at high light levels than their non-crosslinked counterparts. $\square$

Methods

Film preparation

Thin films were prepared by spin-coating the polymers from toluene solution (15 mg ml⁻¹), which additionally contained 0.5 wt% (relative to EL polymer) of the photoacid {4-[(2-hydroxytetradecyl)-oxyl]-phenyl}-phenyliodonium hexafluorantimonate **8** (Fig. 1) as initiator for the cationic polymerization of the oxetane units²¹. The films were irradiated (standard handheld UV lamp, UV B, $\lambda = 302$ nm) for about 3 s in inert gas atmosphere at room temperature. Afterwards, to further advance the crosslinking process in the growing network, the films were annealed at 90 °C for 30 s ('soft curing'). The heating increased the diffusional mobility of the oxetane reactive groups and thus the degree of crosslinking. Afterwards, the films were rinsed with THF and finally heated to 180 °C for 5 min to remove the radical cations formed upon polymerization ('post baking'). The resulting polymer networks were found to be insoluble in any common solvent. The films exhibited an average surface roughness (root mean square, r.m.s.) of <1 nm as determined by atomic-force microscopy (AFM). No signs of shrinkage were found.

Device preparation and physical characterization

ITO (indium tin oxide)-coated glass substrates were commercially obtained from Merck display glass (20 Ω inch⁻²) and carefully cleaned and dried before use. An ozone treatment yielded a rather hydrophilic surface.

Single-colour OLED devices were obtained by first spin-coating PEDOT as the hole-injecting contact (20 nm), baking at 120 °C, followed by spin-coating of the polymers from toluene solution (15 mg ml⁻¹) to yield 80-nm-thin films, then crosslinking the films as described above, and finally depositing Ca (20 nm) and Ag (200 nm) as the top electrode (active area 0.08 cm²). Except for the PEDOT deposition, all other steps were performed in a glove box under inert gas atmosphere.

RGB devices were fabricated by depositing the first polymer as described above, illuminating selected areas through a shadow mask, soft curing, then removing the still soluble, non-illuminated areas with THF. This procedure was repeated for each colour, and finally the devices were post-baked. The results were independent of the sequence of deposition of the three colours, unambiguously demonstrating that a layer once crosslinked is unaffected by the others (no inter-diffusion/penetration).

The characterization of the devices was performed by a custom-made voltage generator using a calibrated amperemeter and a calibrated photo diode.

Chemicals

All chemicals were purchased from Aldrich, Lancaster, or Strem and were used without further purification. Monomers **1** and **3–7** were synthesized as described; see ref. 24. The reactions were carried out under an argon atmosphere. Solvents were commercial pro analysis (p.a.) quality.

Synthesis of 2

Synthesis of 2,7-dibromo-9-(2,5-dimethylphenyl)-9-(3,4-di(3-ethyl(oxetane-3-ethoxy)-hexyloxyphenyl))-fluorene was carried out as follows:

1 g (1.9 mmol) of 2,7-dibromo-9-(2,5-dimethylphenyl)-9-(3,4-dihydroxyphenyl)-fluorene was dissolved in 10 ml dimethyl sulphoxide (DMSO) in a 100-ml Schlenk flask under nitrogen atmosphere. 2.08 g (7.5 mmol, 4 equiv.) 3-(6-bromohexylmethyl)-3-ethyloxetane were added under stirring. The mixture was degassed several times. 0.49 g (7.5 mmol, 4 equiv.) of very fine ground KOH were added. The reaction was performed at 60 °C. After 5 h the mixture was cooled to room temperature and diluted with 20 ml water. The aqueous phase was extracted with ether. The organic phases were collected and dried over anhydrous magnesium sulphate. After removing the solvent the resulting oil was purified by silica-gel column chromatography using toluene/ethylacetate (4:1) as eluent. The product **2** was obtained as a light yellow oil in 67.8% yield.

¹H NMR (CDCl₃) spectrum: 7.66–6.48 (multiplet (m), 12H, aromatic); 4.45–4.30 (m, 8H, -OCH₂-oxetane); 3.95–3.80 (m, 4H, -OCH₂-ether); 3.49 (singlet (s), 2H, -OCH₂-); 3.48 (s, 2H, -OCH₂-); 3.42 (triplet (t), 4H, -OCH₂-); 2.19 (s, 6H, benzylic); 1.82–1.29 (m, 20H, aliphatic); 0.89–0.79 (m, 6H, -CH₃).

General procedure for polymer synthesis

The boronic ester **1** (3.203 g, 4.0 mmol), the mixture of bromide monomers **2–7** (see Table 1; 4.0 mmol in total), and K₃PO₄·H₂O (4.05 g, 17.6 mmol) were added to a mixture of toluene (12 ml), water (6 ml) and ethanol (0.1 ml), and the solution was degassed for 30 min. Pd(PPh₃)₄ (70 mg, 0.06 mmol) was added, and the suspension was heated under reflux (~87 °C) under argon with vigorous stirring. After 2 days boronic ester (0.03 g, 0.058 mmol) was added, and the mixture was refluxed for a further 6 h. Then bromobenzene (0.05 ml, 0.47 mmol) was added for endcapping and the mixture refluxed for a further 3 h. After cooling, the reaction mixture was diluted with toluene (80 ml) and stirred with NaCN (2% in water, 100 ml) for 3 h. The organic phase was separated, washed with water and poured into methanol to precipitate the polymer. This was then dissolved in THF (80 ml) and reprecipitated in methanol (200 ml). The polymer was collected by filtration and dried under vacuum. The later procedure was repeated.

Yield of **P1**: 4.3 g, 77%. ¹H NMR (CDCl₃) spectrum: 7.8–6.6 (m, 22.1H); 6.4–6.0 (m, 2H, Spiro); 4.3–4.4 (m, 4H, OCH₂-oxetane); 4.0–3.3 (2 × m, 15.2H, OCH₂); 2.2–0.7 (m, 61.6 alkyl H).

Yield of **P2**: 2.0 g, 76%. ¹H NMR (CDCl₃) spectrum: 7.8–6.6 (m, 21.8 H); 6.4–6.0 (m, 2H, Spiro); 4.3–4.4 (m, 4H, OCH₂-oxetane); 4.0–3.3 (2 × m, 19.0H, OCH₂); 2.2–0.7 (m, 60.1 alkyl H).

Yield of **P3**: 1.96 g, 75%. ¹H NMR (CDCl₃) spectrum: 7.8–6.6 (m, 19.8H); 6.4–6.0 (m, 2H, Spiro); 4.3–4.4 (m, 4H, OCH₂-oxetane); 4.0–3.3 (2 × m, 14.0H, OCH₂); 2.2–0.7 (m, 55.6 alkyl H).

Yield of **P4**: 3.79 g, 79%. ¹H NMR (CDCl₃) spectrum: 7.8–6.2 (m, 20.4H); 6.2–6.1 (m, 4H, Spiro); 4.0–3.4 (2 × m, 11.2H, OCH₂); 2.1–0.7 (m, 58.8H, alkyl H).

Received 7 October; accepted 31 December 2002; doi:10.1038/nature01390.

1. Friend, R. H. *et al.* Electroluminescence in conjugated polymers. *Nature* **397**, 121–128 (1999).
2. Heeger, A. J. Light emission from semiconducting polymers: light-emitting diodes, light-emitting electrochemical cells, lasers and white light for the future. *Solid State Commun.* **107**, 673–679 (1998).
3. Kraft, A., Grimsdale, A. C. & Holmes, A. B. *Angew. Chem. Int. Edn* **37**, 402–428 (1998).
4. Salbeck, J. Electroluminescence with organic compounds. *Ber. Bunsenges. Phys. Chem.* **100**, 1667–1677 (1996).
5. Pardo, D. A., Jabbar, G. E. & Peyghambarian, N. Application of screen printing in the fabrication of organic light-emitting devices. *Adv. Mater.* **12**, 1249–1252 (2000).
6. Birnstock, J. *et al.* Screen-printed passive matrix displays based on light-emitting polymers. *Appl. Phys. Lett.* **78**, 3905–3907 (2001).
7. Birnstock, J. *et al.* Screen-printed passive matrix displays and multicolor devices. *Proc. SPIE* **4464**, 68–73 (2001).
8. Haskal, E. *et al.* Ink jet printing of passive matrix polymer light-emitting displays. *SID Digest Tech. Pap.* **33**, 776–779 (2002).
9. Seong, T. L. *et al.* A new patterning method for full-colour light-emitting diodes: laser-induced thermal imaging. *SID Digest Tech. Pap.* **33**, 784–787 (2002).
10. Trättnig, C., Pogantsch, A., Langer, G., Kern, W. & Zojer, E. Polymer-based red, green and blue emitting devices fabricated by reductive photopatterning. *Appl. Phys. Lett.* **81**, 4269–4271 (2002).
11. Müller, D. C. *et al.* Efficient blue organic light-emitting diodes with graded hole-transport layers. *ChemPhysChem* **1**, 207–211 (2000).
12. Bayerl, M. S. *et al.* Cross-linkable hole transport materials—Preparation of multilayer organic light-emitting diodes by spincoating. *Macromol. Rapid Commun.* **20**, 224–227 (1999).
13. Li, X. C. *et al.* A blue light emitting copolymer with charge transporting and photo-crosslinkable functional units. *Synth. Met.* **84**, 437–438 (1997).

14. Le Barney, P., Bouche, C. M., Facchetti, H., Soyer, F. & Robin, P. Synthesis of side chain electroluminescent polymers and properties of devices including them. *Proc. SPIE* **3148**, 160–169 (1997).
15. Bellmann, E. *et al.* New triarylamine-containing polymers as hole transport materials in organic light-emitting diodes: effect of polymer structure and crosslinking on device characteristics. *Chem. Mater.* **10**, 1668–1676 (1998).
16. Jiang, X., Liu, S., Ma, H. & Jen, A. K. High-performance blue light-emitting diode based on a binaphthyl-containing polyfluorene. *Appl. Phys. Lett.* **76**, 1813–1815 (2000).
17. Bacher, A. *et al.* Photo-cross-linked triphenylenes as novel insoluble hole transport materials in organic LEDs. *Macromolecules* **32**, 4551–4557 (1999).
18. Li, W. *et al.* Covalently interlinked organic LED transport layers via spin-coating/siloxane condensation. *Adv. Mater.* **11**, 730–734 (1999).
19. Chen, J. P. *et al.* Efficient, blue light-emitting diodes using cross-linked layers of polymeric arylamine and fluorene. *Synth. Met.* **107**, 129–135 (1999).
20. Becker, H., Heun, S., Treacher, K., Büsing, A. & Falcou, A. Materials and inks for full-colour PLED displays. *SID Digest Tech. Pap.* **33**, 780–782 (2002).
21. Fouassier, J.-P. *Photoinitiation, Photopolymerization, and Photocuring: Fundamentals and Applications* 102–144 (Hanser/Gardner, Cincinnati, 1995).
22. Sasaki, H. & Crivello, J. V. The synthesis, characterization, and photoinitiated cationic polymerization of difunctional oxetanes. *J. Macromol. Sci. A* **29**, 915–930 (1992).
23. Penczek, S. & Kubisa, P. in *Comprehensive Polymer Science* (ed. Allen, G.) Vol. III, 751–786 (Pergamon, Oxford, 1989).
24. Treacher, K. *et al.* Conjugated polymers containing spirobifluorene units and fluorene units, and the use thereof. German patent application DE 2001-10114477; WO 2002-077060.

Acknowledgements Partial financial support was granted by the Deutsche Forschungsgemeinschaft, the Bundesministerium für Bildung und Forschung, the Fonds der Chemischen Industrie (Kekule grant for N.R.), and the Bavarian government through 'Neue Werkstoffe'.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.M. (e-mail: klaus.meerholz@uni-koeln.de).

Precise dating of Dansgaard–Oeschger climate oscillations in western Europe from stalagmite data

D. Genty*, D. Blamart*, R. Ouahdi*, M. Gilmour†, A. Baker‡, J. Jouzel* & Sandra Van-Exter§

* IPSL/Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA/CNRS 1572 Bat. 709, L'Orme des Merisiers CEA Saclay, 91191 Gif sur Yvette cedex, France

† The Open University, Department of Earth Sciences, Uranium Series Facility, Milton Keynes MK7 6AA, UK

‡ Centre for Land Use and Water Resources Research, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

§ University of Montpellier, Laboratoire Hydrosociétés, UMR 5569 (CNRS-UM2-IRD), Maison des sciences de l'eau, CC 057, Place E. Bataillon, 34095 Montpellier cedex 5, France

The signature of Dansgaard–Oeschger events—millennial-scale abrupt climate oscillations during the last glacial period—is well established in ice cores and marine records^{1–3}. But the effects of such events in continental settings are not as clear, and their absolute chronology is uncertain beyond the limit of ¹⁴C dating and annual layer counting for marine records and ice cores, respectively. Here we present carbon and oxygen isotope records from a stalagmite collected in southwest France which have been precisely dated using ²³⁴U/²³⁰Th ratios. We find rapid climate oscillations coincident with the established Dansgaard–Oeschger events between 83,000 and 32,000 years ago in both isotope records. The oxygen isotope signature is similar to a record from Soreq cave, Israel⁴, and deep-sea records^{5,6}, indicating the large spatial scale of the climate oscillations. The signal in the carbon isotopes gives evidence of drastic and rapid vegetation

changes in western Europe, an important site in human cultural evolution. We also find evidence for a long phase of extremely cold climate in southwest France between 61.2 ± 0.6 and 67.4 ± 0.9 kyr ago.

Several of the 24 Dansgaard–Oeschger (DO) events identified in the GRIP^{7,8} and GISP2⁹ Greenland ice cores over the last glacial period indicate air temperature changes greater than 10 °C. North Atlantic and Mediterranean marine cores show similar amplitude changes in the water temperature^{5,10,11}. However, the consequences of these events on the continent are still not well known, because of the scarcity of well-preserved and well-dated records. We know that the ocean–atmosphere system was closely coupled as far as the central Mediterranean region, where the vegetation has been reconstructed thanks to pollen records (in lakes) that reacted very quickly and sensitively to these abrupt climate changes^{12,13}. GRIP and GISP2 dating uncertainties may exceed the duration of a single DO event for a large part of the glacial period (estimated for GISP2 at 2% over the past 40 kyr, and between 5% and 10% for the period <50 kyr ago^{14,15}). Marine cores are dated by orbital tuning with inherent accuracy and, for their late-glacial part, by ¹⁴C dating. However, for periods of interest here, this latter method suffers from limitations linked with analytical precision, ¹⁴C age calibration, fluctuations in the carbon cycle, and a varying age reservoir¹⁶. The few continental records also display a lack of absolute ages, especially after 40 kyr ago^{12,13}. This highlights the importance of establishing chronologies of those events from other records.

Here we present stable isotope records ($\delta^{13}\text{C}$ and $\delta^{18}\text{O}$) from a stalagmite of the Villars cave (southwest France, less than 200 km from the Atlantic Ocean; Fig. 1) that show most of the DO events between 83 and 32 kyr ago (Fig. 2). We have determined 27 thermal ionization mass spectrometry (TIMS) U-series dates on this 147-cm-long stalagmite ('Vil9'), leading to an absolute timescale with errors less than 2% (2σ) up to 83 kyr ago (see Supplementary Table 1 for information, and ref. 17 for methods). Three main hiatuses occurred at 78.8–75.5 kyr ago (referred to as D2), 67.4–61.2 kyr ago (D3) and 55.7–51.8 kyr ago (D4). The regular $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ decreases towards the D3 hiatus, and also the steady increase that is observed afterwards, suggests that this growth cessation is

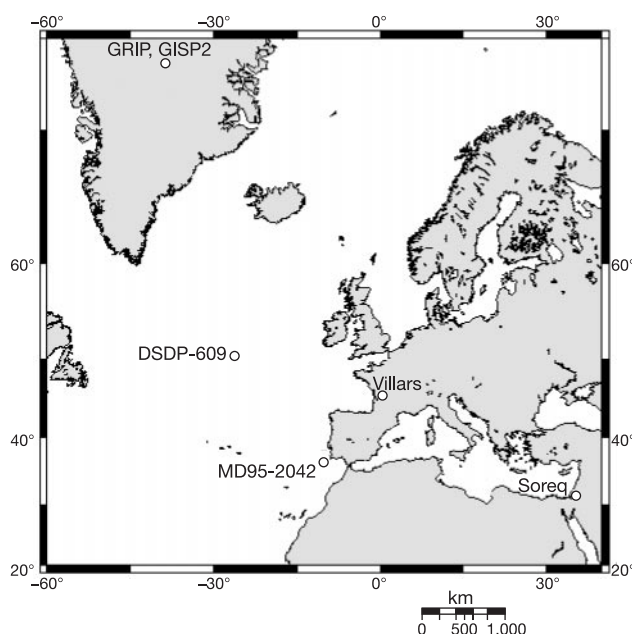


Figure 1 Map showing locations of the study sites. GRIP and GISP2; 72.58° N, 37.64° W. Villars cave; 45.30° N, 0.50° E. MD95-2042; 37.48° N, 10.10° W. DSDP 609; 50° N, 27° W. Soreq cave; 31.6° N, 35° E.

related to climatic and/or vegetation causes. The timing of D3 is in good temporal agreement with Heinrich event H6, whose duration, from other records, is not well constrained⁵. This confirms that the climate switched to much colder conditions leading to the cessation of stalagmite growth for about 6.2 kyr, which is one of the main features of the Vil9 record ('Villars cold phase', Fig. 3). Because D2 and D4 hiatuses occurred during stable or low $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ levels, their cause is possibly different from that of D3 (perhaps flooding). The other visible discontinuities, where no growth hiatus occurred, are due either to very short cessations, or to short events that

brought detrital particles onto the stalagmite growth surface (such as flooding or micro-fissure emptying). The fact that the other Heinrich events (that is, H4, H5) are not recorded on our sample as growth cessations suggests that they led to less-drastring climatic conditions in southwest France, as is also suggested by nearby marine cores DSDP-609 and MD95-2042 (Figs 3 and 4).

DO events are particularly visible in the carbon isotope record, with typically 2–4‰ variations in the $\delta^{13}\text{C}$. A comparison with $\delta^{18}\text{O}$ GRIP and GISP2 records allows their identification, low $\delta^{13}\text{C}$ values corresponding to high ice $\delta^{18}\text{O}$ and thus to warmer conditions and,

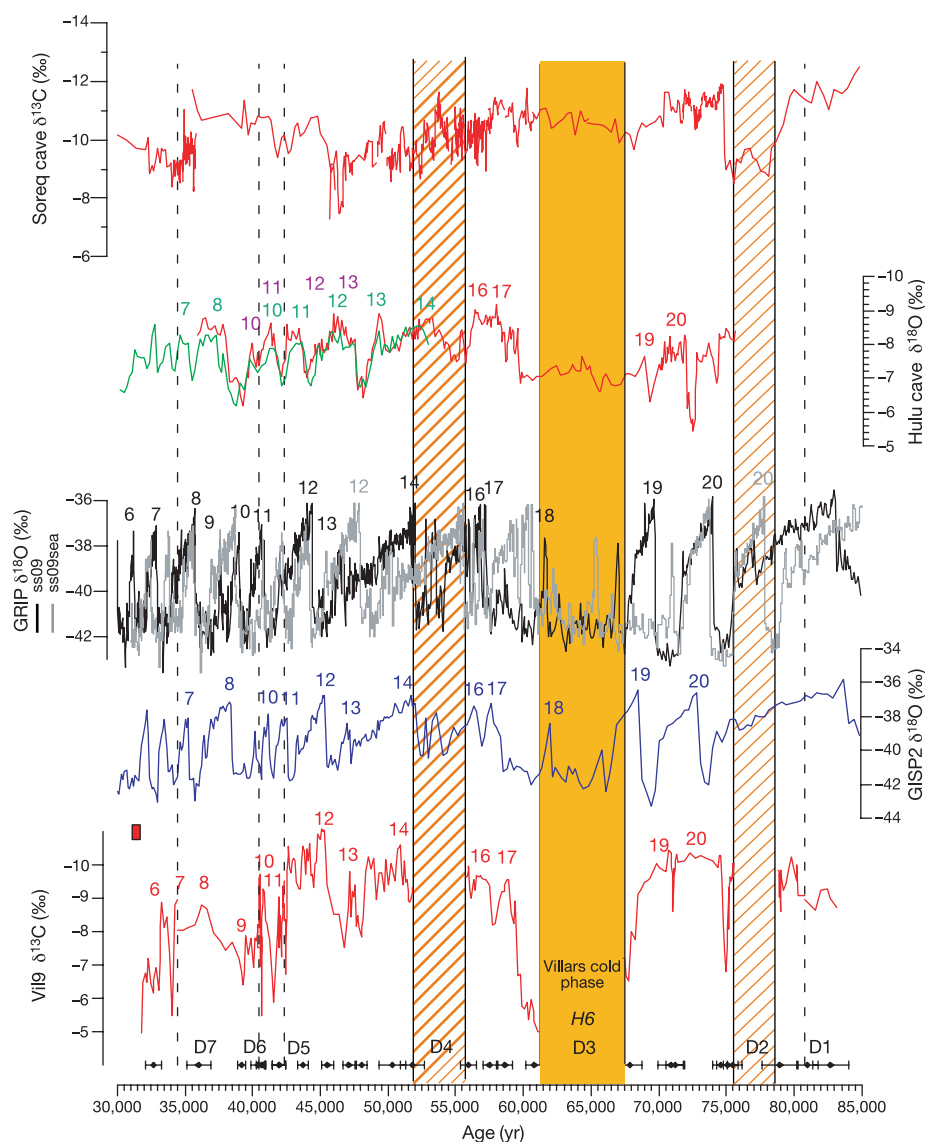


Figure 2 Comparison of the Vil9 stalagmite carbon isotope profile with Greenland ice core records and other stalagmite records. Vil9 stalagmite $\delta^{13}\text{C}$ record compared with GISP2 ice core $\delta^{18}\text{O}$ (ref. 9), GRIP ice core $\delta^{18}\text{O}$ (ss09sea chronology; ref. 8), GRIP ice core $\delta^{18}\text{O}$ (ss09sea chronology; ref. 24). Note that DO events from GRIP ss09sea chronology are all significantly older than in Vil9 DO with a lag of about one entire event. Also shown are records from Hulu cave (China) stalagmite $\delta^{18}\text{O}$ (ref. 18) and from Soreq cave, Israel ($\delta^{13}\text{C}$) (ref. 4). The fact that the Soreq cave $\delta^{13}\text{C}$ record does not show the same features can be explained either by difference in local hydrology and by local vegetation influences (that is, C_4 type) or because it is too far from the Atlantic. At the bottom of the figure, D1–D7 indicate visible discontinuities, and H6 shows a Heinrich event. Modern calcite $\delta^{13}\text{C}$ values at Villars correspond to the red rectangle at the bottom left. Our chronology, like all others that use absolute U-series ages, might have been affected by an 'accordion'

effect because growth rate is likely to have changed with climate: growth rate is 0.1 mm yr^{-1} during DO 12, one of the warmest events, whereas it is 0.01 mm yr^{-1} at the end of the stalagmite growth, the coldest period which led to the definitive stop of the growth. Speleothem growth rate is mainly controlled by the seepage water calcium concentration and by the drip rate²⁵. In the studied area, and for the present day, it varies from 0.3 mm yr^{-1} to about 1.0 mm yr^{-1} (ref. 26). It is likely that during cold periods, the calcium concentration in the seepage water was lower, owing to lower vegetation activity, leading to a slower growth rate. This might have happened between each warm phase of DO events; however, we have dated points that surround these cold phases (between DO 19 and 20, before DO 20, between DO 10 and 11), and it appears that their durations are not particularly long (that is, less than 1 kyr). Coloured vertical bars represent growth hiatuses.

at least in Greenland, to higher precipitation (Fig. 2). Most of GISP2 DO ages are within the 2σ error limits, except for DO 8, 19 and 20. DO 8 displays a regular $\delta^{13}\text{C}$ decrease from 39.4 to 36.2 kyr ago whose shape is very similar to the Hulu cave record¹⁸. DO 19 and 20 events are significantly older in our record, and are better correlated with the GRIP (ss09) record. The duration of the cold phases surrounding these two former events seems to have been very short, as shown by the four U/Th ages that have been determined just at the edge of their limits (Fig. 2). DO 12 reached its optimum (lower $\delta^{13}\text{C}$) at 45.3 ± 0.4 kyr ago, similar to the GISP2 chronology and about 1,000 yr older than in GRIP ss09 chronology, but it seems to have started much earlier at about 46.8 ± 0.4 kyr ago. DO 12 is also the most pronounced event, as determined by its stable isotope amplitude variation, by its duration and by the stalagmite growth rate: between 42.3 and 45.5 kyr ago, 16% of the Vil9 height grew for less than 6% of its growth duration. DO 12 is characterized by a well-marked saw-tooth shape, very similar to the GRIP record with a $\delta^{13}\text{C}$ decrease (climate amelioration) that is clearly visible at the end of the event, just before its abrupt termination at 42.3 kyr ago (Fig. 2). DO events 6 to 14 have been enumerated by reference to DO 12, but we note that their chronology is less clear between DO 8 and 12. The Hulu cave record¹⁸ agrees well with our results if we consider the second chronology (Hulu cave 2), but DO 12 still seems too young (by 1.5 kyr) or too old (by 2.2 kyr) depending on the adopted chronology. We note also that DO 19 is much more pronounced in the Vil9 record.

Considering that most of the record was deposited under isotopic equilibrium (Supplementary Information), the fact that the DO events are so clearly recorded in the $\delta^{13}\text{C}$ profile may be explained as follows. First, a change in the ratio of C_4 plants to C_3 plants suggested in other studies¹⁹ cannot be invoked here, as, until now,

no evidence of C_4 plants have been found in this area during the last glacial. Second, we have shown that most of the carbon (80–90%) of the stalagmites, at least in Villars, comes from the soil CO_2 , and that any change in biologic soil activity conditions will change the carbon isotope signature of these stalagmites¹⁷. Consequently, degradation of vegetation and soil will decrease the biogenic CO_2 production and increase the relative proportion of atmospheric CO_2 , establishing a direct link with temperature and precipitation changes associated with DO events. Two observations support this interpretation: first the $\delta^{13}\text{C}$ of present-day calcite is much higher in caves above which the soil is poor²⁰; second, $\delta^{13}\text{C}$ records of the last glacial–interglacial transition in a stalagmite from southern France show much higher $\delta^{13}\text{C}$ values (that is, $>5\text{‰}$) during the late glacial than during the Holocene (D.G. *et al.*, unpublished results).

Here we suggest that temperature and humidity were important in the development of local vegetation, and consequently in the production of soil CO_2 : Vil9 $\delta^{13}\text{C}$ is lower during warm periods because a higher proportion of biogenic CO_2 is dissolved in the seepage water. Clear similarities with the MD-95-2042 $\delta^{18}\text{O}$ planktonic record⁶ and with the DSDP-609 foraminifera record⁵, both close to the Villars cave (Fig. 1), support this interpretation (Fig. 3). The time lag between the climate amelioration and the vegetation change seems to have been relatively short for most of the observed DO events, as suggested by the rapid $\delta^{13}\text{C}$ changes observed and also by the good synchronicity between $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ records. This can be explained by a relatively good preservation of the soils during cold periods permitting a fast start to re-vegetation. A record from Monticchio lake has shown a similarly rapid reaction of vegetation to rapid temperature changes¹³. On the contrary, a significant time delay of 1–2 kyr seems to have occurred after the ‘Villars cold phase’,

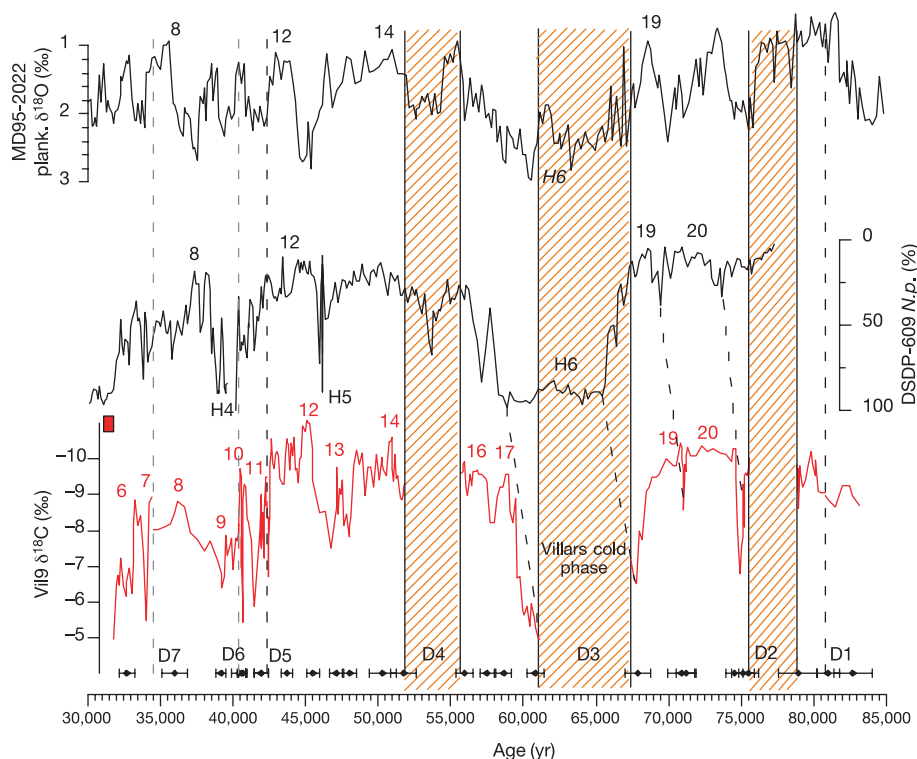


Figure 3 Vil9 $\delta^{13}\text{C}$ record compared with the percentage of *Neoglobigerina pachyderma* of the marine core DSDP 6095, and with planktonic foraminifera $\delta^{18}\text{O}$ of core MD95-2042 (southwest Portugal)⁶. Shape similarities between Vil9 and DSDP 609 records are particularly striking for the oldest part of the record (>50 kyr ago), highlighting the occurrence of a long cold period between 61.2 ± 0.59 and

67.4 ± 0.87 kyr ago that we called the ‘Villars cold phase’. It suggests that the local temperature was very similar to the regional SST, imprinted in the marine core records, and followed the warming–cooling events. Coloured vertical bars represent growth hiatuses.

as shown by the long and regular $\delta^{13}\text{C}$ decrease demonstrating the harshness of this period (Fig. 2). Finally, temperature variation might also have played a significant role in the $\delta^{13}\text{C}$ changes: equilibrium fractionation during soil CO_2 dissolution and during calcite precipitation in the cave, leading to an average $\delta^{13}\text{C}$ gradient of -0.1‰ per $^\circ\text{C}$ (ref. 21), might explain about 20% of the observed $\delta^{13}\text{C}$ variations.

On the Vil9 stalagmite, the sequence of DO events is less clearly depicted in the $\delta^{18}\text{O}$ than in the $\delta^{13}\text{C}$ record (Fig. 4). In addition to the well-marked DO 12 event, the $\delta^{18}\text{O}$ record shows significant general increasing trends between 75.5 and 67.4 kyr ago ($+2.1\text{‰}$) and between 51.8 to 31.8 kyr ago ($+1.4\text{‰}$). Those trends parallel the SPECMAP and the benthic MD-95-2042 records, leading to the hypothesis that they result from the deterioration of the global climate. Superimposed on these general features, the Vil9 $\delta^{18}\text{O}$ record shows changes associated with DO events probably linked with the temperature changes (Fig. 4).

An estimation of the temperature (Fig. 4) shows a drop between 75 and 67.4 kyr ago of $-13.6 \pm 2^\circ\text{C}$, which, compared to the present-day temperature ($12.5 \pm 0.5^\circ\text{C}$), suggests that the local temperature was close to freezing just before the 6.2-kyr-long

'Villars cold phase'. This period fits well with the GRIP ss09 record where ice $\delta^{18}\text{O}$ stayed generally low (-42‰), and was punctuated by two warming events (DO 18 and another one) that were too short to start the seepage water process again above the Villars cave. We suppose that during this non-deposition period (D3) a permafrost occurred above the Villars cave. The stalagmite stopped growing after a general cooling trend punctuated by nine DO events and marked, at the end, by the lowest growth rate (0.01 mm yr^{-1} between 38.9 and 31.9 kyr ago). During this last cooling trend, DO 12 is the most pronounced event, characterized by the highest growth rate and the highest $\delta^{18}\text{O}$ variation ($>1.2\text{‰}$) between the coldest and the warmest phase. Applying the same assumptions as for the 75.5–67.4 kyr trend period implies a temperature increase of more than 10°C , which is in good agreement with marine cores¹¹ ($\sim 10^\circ\text{C}$) and pollen records from lakes in southern and eastern France¹² ($\sim 8^\circ\text{C}$).

Southwest France (and more generally, southwest Europe) possesses many caves that, given their southern location away from the British Isles, Ireland and Fennoscandian ice sheets, surely recorded DO events in the same way as the Vil9 stalagmite. The wide availability of TIMS uranium dates permits precise chronologies

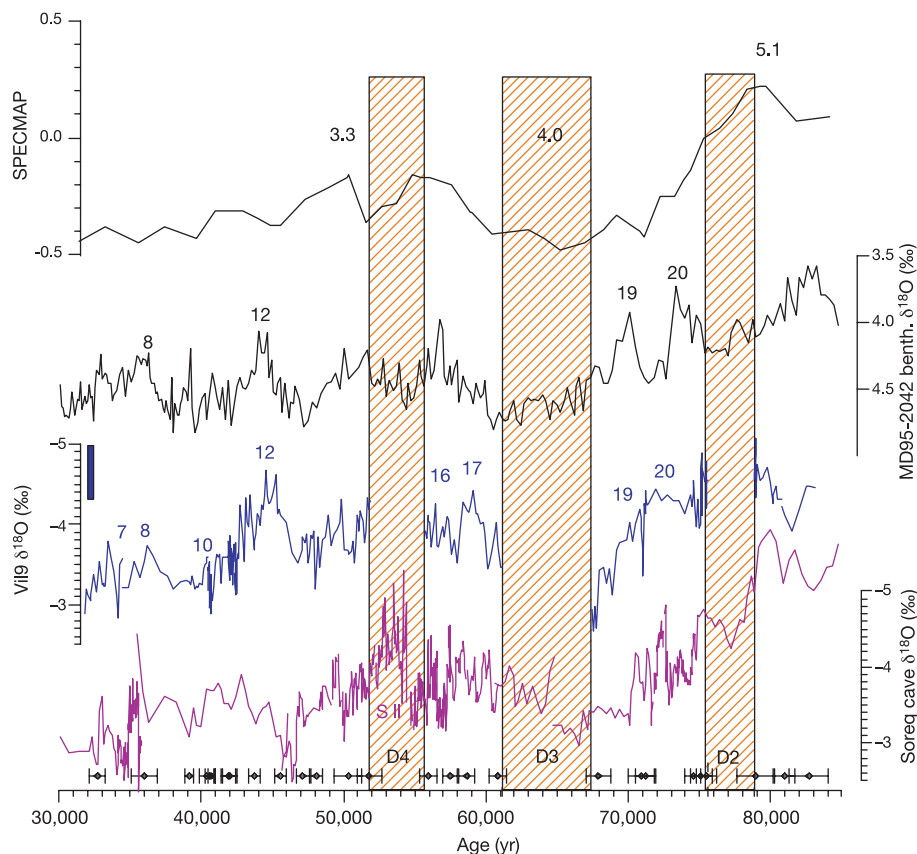


Figure 4 Comparison of the Vil9 $\delta^{18}\text{O}$ record with the Soreq cave $\delta^{18}\text{O}$ record⁴, benthic foraminifera $\delta^{18}\text{O}$ of marine core MD95-2042⁶, and SPECMAP $\delta^{18}\text{O}$ normalized record with MIS numbers. Modern calcite $\delta^{18}\text{O}$ values at Villars are shown by the rectangle on the left. There is a marked resemblance between Vil9 and Soreq cave $\delta^{18}\text{O}$ records, not only in the general trends but also in the absolute $\delta^{18}\text{O}$ values. This suggests that both records are affected in a similar way by the general climate deterioration associated with the growth of continental ice sheets. Small differences, for example the absence of the DO 12 event in the Soreq record, or the absence of Sapropel 2 in Villars record (D4), are the consequences of climate regional influences and/or local geomorphology factors which allow or prevent flooding in the cave. As pointed out by many authors^{4,18,19,27}, the speleothem $\delta^{18}\text{O}$ signal is influenced by many parameters: the seepage water $\delta^{18}\text{O}$ ($\delta^{18}\text{O}_w$), the ocean $\delta^{18}\text{O}$, the depression path, the local temperature, the seasonality of the

precipitation and finally the calcite deposition conditions (temperature, isotopic equilibrium), where temperature-dependent fractionation during calcite precipitation acts in the opposite way to external temperature (about -0.23‰ per $^\circ\text{C}$; ref. 28). Theoretically, calculation of past temperatures is possible if $\delta^{18}\text{O}_w$ can be estimated or, better, deduced from fluid inclusions trapped in the calcite lattice²⁹. Measurements of fluid inclusions are under development in our laboratory; here we tentatively follow the first approach in assuming that the change of $\delta^{18}\text{O}_w$ for the coldest period (67.4 kyr ago) with respect to its present-day value is close to the change estimated for the last glacial over western Europe ($-1.5 \pm 0.5\text{‰}$; ref. 30), and that during the warmest phase of this period, at 75 kyr ago, $\delta^{18}\text{O}_w$ was close to its present-day value (average seepage water $\delta^{18}\text{O}$ measured in Villars cave is $-6.3 \pm 0.1\text{‰}$).

for climate events to be developed for the past 400 kyr. This constrains not only ice-core chronologies, but also all other records that are dependent upon ice-core timescales. For example, a simple intercomparison between VIL9 and GRIP records permits the precise dating of the late glacial ^{10}Be peak, which largely covers DO 10 (ref. 22), an event extending between 40.4 and 41.5 kyr ago in the VIL9 record. Finally, it is likely that such large climatic oscillations have affected past human cultures, causing migrations and changes in living and eating habits. Archaeological remains, which can be time constrained thanks to calcite deposits found in prehistoric shelters, can then be put in a climatic framework that will lead to a better understanding of these past cultures²³. □

Received 27 September; accepted 23 December 2002; doi:10.1038/nature01391.

- Labeyrie, L. Glacial climate instability. *Science* **290**, 1905–1907 (2000).
- Blunier, Th. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
- Bond, G. C. *et al.* Persistent solar influence on North Atlantic climate during the Holocene. *Science* **294**, 2130–2136 (2001).
- Bar-Matthews, M., Ayalon, A. & Kaufmann, A. Timing and hydrological conditions of sapropel events in the Eastern Mediterranean, as evidence from speleothems, Soreq Cave. *Chem. Geol.* **169**, 145–156 (2000).
- Bond, G. *et al.* Correlations between climate records from North Atlantic sediment and Greenland ice. *Nature* **365**, 143–147 (1993).
- Shackleton, N. J. & Hall, M. A. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15**, 565–569 (2001).
- Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Groote, P. M., Stuiver, M., White, J. W. C., Johnsen, S. J. & Jouzel, J. Comparison of the oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
- Cacho, I. *et al.* Dansgaard-Oeschger and Heinrich event imprints in Alboran Sea paleotemperatures. *Paleoceanography* **14**, 698–705 (1999).
- Cayre, O., Lancelot, Y., Vincent, E. & Hall, M. A. Paleoclimatological reconstructions from planktonic foraminifera of the Iberian margin: Temperature, salinity, and Heinrich events. *Paleoceanography* **14**, 384–396 (1999).
- Guiot, J. *et al.* The climate in Western Europe during the last glacial/interglacial cycle derived from pollen and insect remains. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **103**, 73–93 (1993).
- Allen, J. R. M. *et al.* Rapid environmental changes in southern Europe during the last glacial period. *Nature* **400**, 740–743 (1999).
- Bender, M. *et al.* Climate connection between Greenland and Antarctica during the last 100,000 years. *Nature* **372**, 663–666 (1994).
- Lang, C. *et al.* 16 °C rapid temperature variation in Central Greenland 70000 years ago. *Science* **286**, 934–937 (1999).
- Beck, W. J. *et al.* Extremely large variations of atmospheric ^{14}C concentration during the last glacial period. *Science* **292**, 2453–2458 (2001).
- Genty, D. *et al.* Dead carbon in stalagmites: limestone paleodissolution versus ageing of soil organic matter—Implications for ^{13}C variations in stalagmites. *Geochim. Cosmochim. Acta* **65**, 3443–3457 (2001).
- Wang, Y. J. *et al.* A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
- Dorale, J. A., Edwards, R. L., Ito, E. & Gonzalez, L. Climate and vegetation history of the midcontinent from 75 to 24 ka: A speleothem record from Crevice Cave, Missouri, USA. *Science* **282**, 1871–1874 (1998).
- Urbanc, J., Pezdirc, J., Dolenec, T. & Perko, S. Isotopic composition of oxygen and carbon in cave water and speleothems of Slovenia. *Acta Carsologica* **13**(4), 99–112 (1985) (in Slovenian).
- Mook, W. G. in *Handbook of Environmental Geochemistry* (eds Fritz, P. & Fontes, J. Ch.) Vol. 1-A, 49–74 (Elsevier Science, Amsterdam, 1980).
- Yiou, F. *et al.* Beryllium 10 in the Greenland Ice Core Project ice core at Summit Greenland. *J. Geophys. Res.* **102**, 26783–26794 (1997).
- Stringer, C. & Davies, W. Those elusive Neanderthals. *Nature* **413**, 791–792 (2001).
- Johnsen, S. J. *et al.* Oxygen isotope and paleotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and North GRIP. *J. Quat. Sci.* **16**, 299–307 (2001).
- Baker, A., Smart, P. L., Edwards, R. L. & Richards, D. A. Annual growth bandings in a cave stalagmite. *Nature* **364**, 518–520 (1993).
- Genty, D. & Quinif, Y. Annually laminated sequences in the internal structure of some Belgian stalagmites—Importance for paleoclimatology. *J. Sedim. Res.* **66**, 275–288 (1996).
- Hendy, C. H. The isotopic geochemistry of speleothems—I. The calculation of the effects of different modes of formation on the isotopic composition of speleothems and their applicability as paleoclimatic indicators. *Geochim. Cosmochim. Acta* **35**, 801–824 (1971).
- O’Neil, J. R., Clayton, R. N. & Mayeda, T. K. Oxygen isotope fractionation in divalent metal carbonates. *J. Chem. Phys.* **51**, 5547–5558 (1969).
- Dennis, P. F., Rowe, P. J. & Atkinson, T. C. The recovery and isotopic measurement of water from fluid inclusions in speleothems. *Geochim. Cosmochim. Acta* **65**, 871–884 (2001).
- Rozanski, K. Deuterium and oxygen-18 in European groundwaters—Links to atmospheric circulation in the past. *Chem. Geol.* **52**, 349–363 (1985).

Supplementary Information accompanies the paper on *Nature’s* website
(<http://www.nature.com/nature>).

Acknowledgements We thank H. Versaveau and Th. Baritaud for help in sampling, and C. Waelbroeck, U. von Grafenstein, G. Hoffmann, D. Joly and M. F. Diot for discussions. This work was supported by NERC, CNRS–INSU (VariEnTe, Eclipse, PNEDC) and the CEA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.G.
(e-mail: genty@lsce.saclay.cea.fr).

New ages for human occupation and climatic change at Lake Mungo, Australia

James M. Bowler^{*†}, Harvey Johnston[‡], Jon M. Olley[§], John R. Prescott^{||}, Richard G. Roberts^{†¶}, Wilfred Shawcross[#] & Nigel A. Spooner^{†☆☆}

^{*} School of Earth Sciences, University of Melbourne, Melbourne, Victoria 3010, Australia

[‡] NSW National Parks & Wildlife Service, Buronga, New South Wales 2739, Australia

[§] CSIRO Land & Water, Canberra, Australian Capital Territory 2601, Australia

^{||} Department of Physics & Mathematical Physics, University of Adelaide, Adelaide, South Australia 5005, Australia

[¶] School of Geosciences, University of Wollongong, Wollongong, New South Wales 2522, Australia

[#] 25 Fairfax Street, O’Connor, Canberra, Australian Capital Territory 2602, Australia

[☆] Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 0200, Australia

[†] These authors contributed equally to this work

Australia’s oldest human remains, found at Lake Mungo, include the world’s oldest ritual ochre burial (Mungo III)¹ and the first recorded cremation (Mungo I)². Until now, the importance of these finds has been constrained by limited chronologies and palaeoenvironmental information³. Mungo III, the source of the world’s oldest human mitochondrial DNA⁴, has been variously estimated at 30 thousand years (kyr) old¹, 42–45 kyr old^{5,6} and 62 ± 6 kyr old^{7,8}, while radiocarbon estimates placed the Mungo I cremation near 20–26 kyr ago^{2,9,10}. Here we report a new series of 25 optical ages showing that both burials occurred at 40 ± 2 kyr ago and that humans were present at Lake Mungo by 50–46 kyr ago, synchronously with, or soon after, initial occupation of northern^{11,12} and western Australia¹³. Stratigraphic evidence indicates fluctuations between lake-full and drier conditions from 50 to 40 kyr ago, simultaneously with increased dust deposition, human arrival and continent-wide extinction of the megafauna^{14,15}. This was followed by sustained aridity between 40 and 30 kyr ago. This new chronology corrects previous estimates for human burials at this important site and provides a new picture of *Homo sapiens* adapting to deteriorating climate in the world’s driest inhabited continent.

Lake Mungo, the centrepiece of the Willandra Lakes World Heritage Area in western New South Wales, boasts several features of significance to world archaeology, such as Quaternary climate change and the interaction of early humans with the Australian landscape and biota. It is also the type site for the Lake Mungo Geomagnetic Excursion¹⁶.

Fed by an ancestral course of the Lachlan River rising in the highland catchments near Canberra (Fig. 1), the hydrologic record from the now dry lake basin reflects Pleistocene changes in

^{**} Present address: DSTO, Edinburgh, South Australia 5111, Australia.

temperature, precipitation and runoff from the then cold, periglacial catchments³. Located in the zone of prevailing westerly winds, the basin has been influenced both by local dunefield activity and by dust from the central Australian dunefields further west. Severe erosion of the large crescent-shaped dune (lunette) on the eastern and southern shores of Lake Mungo has revealed its internal stratigraphy and uncovered substantial quantities of archaeological material (including burials) at the surface.

A basal lake-shore sand (Lower Mungo unit) deposited under lake-full conditions contains a rich assemblage of archaeological sites. The discovery in that unit of burnt human remains (Mungo I)² in 1969 was followed soon after (1974) by the discovery and excavation of a fully articulated skeleton (Mungo III) within the

same stratigraphic unit¹, only 450 m from the cremation site. Artefacts from deep in a trench (Mungo B) near the cremation site remained undated¹⁷. Initial cremation age estimates, based on radiocarbon dating, provided minimum values of 26 kyr (ref. 2), while thermoluminescence ages of 43 ± 4 and 41 ± 7 kyr were obtained later from sands collected at the Mungo III burial site^{5,6}.

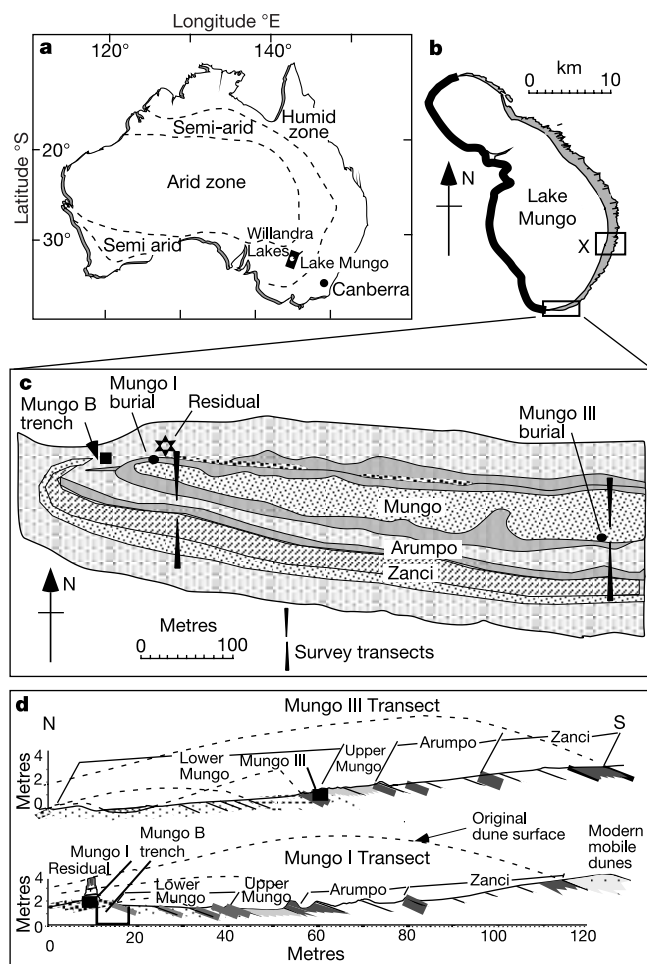


Figure 1 Location of study area. **a**, Location diagram showing Lake Mungo, the centrepiece of the Willandra Lakes World Heritage Area, in semi-arid southeastern Australia. **b**, Lake Mungo, a dry basin last filled about 22,000 years ago, has a large transverse dune (lunette) on its eastern (downwind) margin. The site of the Lake Mungo Geomagnetic Excursion¹⁶ is shown at X, some 8 km north of the burial sites. **c**, The southern sector of the Lake Mungo lunette has been truncated to a nearly horizontal surface, exposing in plan the internal dune stratigraphy, from which a geological map of the eroded surface has been constructed. **d**, Two surveyed transects across the Lake Mungo lunette at the burial sites of Mungo I and III illustrate the depth of erosion, exposing related soil-sedimentary units which can be correlated between sites. Sampling along both transects for optical dating and sedimentary analyses was supplemented by vertical sampling of the Mungo I Residual (bottom left) and the adjacent archaeological excavation (Mungo B trench). The two main units (Lower and Upper Mungo) described here form the inner core of the dune, which contains evidence of human occupation. The Mungo I and III burial sites are located on the northern (lake-shore) and southern (lee) sides of the dune axis, respectively.

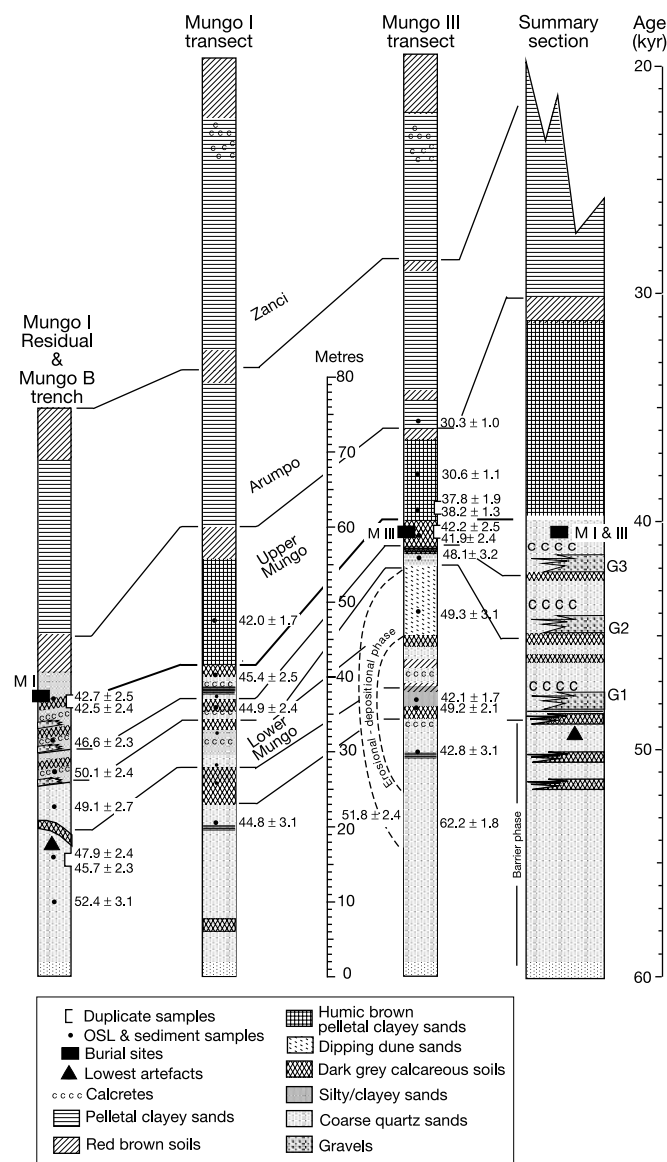


Figure 2 Stratigraphic units and age ranges of sites. The three stratigraphic columns on the left represent the measured and dated sequences. Optical ages are shown in thousands of years (kyr). Data from the Mungo I Residual and Mungo B trench are superimposed vertically for clarity of presentation. Three gravel units (G1, G2, G3) overlie barrier sands, with a band of dark humic, weak soil above the lowest recorded artefacts. The Mungo I and III transects are presented in vertical orientation to illustrate the relative unit thicknesses and to facilitate correlations between surveyed sections. The metric scale shows the actual distances from zero datum (see Fig. 1). The Mungo I and III burials (MI and MIII) have bracketing optical ages that are concordant and also consistent with stratigraphic correlations between the two transects. In the erosional cut-and-fill segment on the Mungo III transect, a unit of unconformable silty sands at 21.0 m (dated to 51.8 ± 2.4 kyr) is stratigraphically equivalent to steeply dipping dune sands near 48 m (49.3 ± 3.1 kyr; see stratigraphic reconstruction in Fig. 1). The synthetic section on the right summarizes the major units and sub-units for all sequences on the left, plotted alongside a timescale determined from the optically stimulated luminescence (OSL) chronology.

Recently, direct dating of the Mungo III skeletal remains produced Th/U and Pa/U ages of between 50.7 ± 0.9 and 82 ± 7 kyr, and electron-spin resonance (ESR) ages from a fragment of tooth enamel of 63 ± 6 (early U-uptake) and 78 ± 7 kyr (linear U-uptake)^{7,8}. Optical ages of 59 ± 3 and 63 ± 3 kyr were obtained for sands collected approximately 300 m away from the burial site and were then thought to relate to the sediments into which Mungo III was inserted⁸. The reliability of these age estimates has been contested^{18–20}.

In this study, improved resolution for the time interval beyond the range of radiocarbon dating has been achieved by optical dating of the lunette sediments, using the optically stimulated luminescence (OSL) signal from quartz^{21,22}. Our multidisciplinary

approach, involving four separate dating laboratories in both field analysis and laboratory studies, provides the first well-constrained ages for the key archaeological and palaeoenvironmental events at Lake Mungo.

Stratigraphic surveys were made of units that range in age from the last interglacial (culturally sterile Golgol unit), through the critical interval from 60 to 40 kyr ago (Lower Mungo unit), to the Last Glacial Maximum (Upper Mungo, Arumpo and Zanci units)³. Attention was focused on two north–south transects across the near-horizontal surface of the eroding lunette: one at the Mungo I burial site and the Mungo B trench, and the other at the Mungo III burial site, 450 m further east (Fig. 1). Survey profiles tied to local benchmarks provide control of horizontal and vertical co-ordinates for all samples and stratigraphic unit boundaries. Stratigraphic units identified in the horizontal transects are presented as vertical sections (Fig. 2), together with dated sample locations and ages. Correlation between transects permits establishment of a synthetic stratigraphic column showing the major hydrologic changes with the main archaeological and megafaunal features (Fig. 3). Mega-faunal remains have been discovered near Lake Mungo²³ and in the adjacent regions^{14,15}, but none were dated in this study.

Sediment samples for optical dating were collected mainly from the Lower Mungo unit, together with corresponding *in situ* gamma-ray spectrometry measurements. Duplicate samples were collected for sediment analyses, with particular attention to microscopic fabric and dust index estimates (Fig. 3d). These record the number of red, clay-coated, silt-sized quartz grains (Wüstenquarz²⁴), which were measured in thin-sections of undisturbed sediment samples on a petrographic microscope (see Supplementary Information). OSL analyses were carried out on sand-sized quartz grains extracted from the Mungo III transect samples (N. A. Spooner), and from the Mungo I transect, the adjacent sediment 'Residual' and the Mungo B trench (R. G. Roberts); four samples were duplicated by both laboratories. Optical ages are shown in their stratigraphic context in Fig. 2.

The optical ages for the Mungo III transect are in correct stratigraphic order, and range from about 62 to 30 kyr ago. Ages for the Mungo I transect, the Residual and the trench (Fig. 2) illustrate sequential development of the Lower Mungo unit from beyond 50 kyr to near 40 kyr ago. Confidence in the reliability of this new chronology for Lake Mungo is given by the stratigraphic correlations and consistency of ages between the two transects; by the agreement between the ages of the four duplicate samples (Fig. 2); and by the internally consistent results obtained for samples for which equivalent doses and dose rates were determined using multiple methods (see Supplementary Information). A minor anomaly occurs where ages obtained for clean quartz sands from near the top of the barrier phase³ are consistently younger than the ages for the stratigraphically higher levels (Mungo B trench at 2.40 m: 45.7 and 47.9 kyr; Mungo I transect at 20.2 m: 44.8 kyr; and Mungo III transect at 29.4 m: 42.8 kyr). We attribute this anomaly to uncertainties in the dose rate, which has a relatively greater effect on the accuracy of the optical ages of these samples because of their very low dose rates.

The first evidence for human occupation is provided by 11 silcrete flakes with plain and relatively thick striking platforms recovered from below the lowest gravels in the barrier sands of the Mungo B trench¹⁷ (Fig. 3b) and bracketed by ages of 50.1 ± 2.4 and 45.7 ± 2.3 kyr. The underlying deposits, dated to 52.4 ± 3.1 kyr ago, appear to be culturally sterile. These ages are similar to, or slightly younger than, those for initial human arrival in northern and western Australia^{11–13}, confirming completion of continental colonization soon after 50 kyr ago. In addition to charcoal and burnt bone, the Mungo B excavation¹⁷ yielded 775 artefacts, which indicate a pattern of discontinuous occupation (Fig. 3b). Maximum occupational density occurred between 45 and 43 kyr ago, preceding the Lower to Upper Mungo transition. The

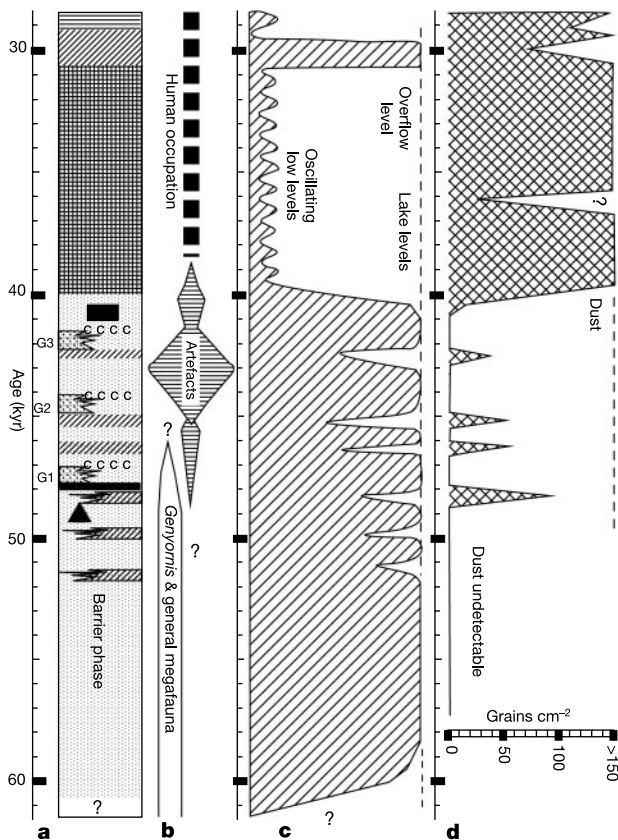


Figure 3 Major hydrological changes and main archaeological and megafaunal features of sites. **a**, Synthetic stratigraphic section (from Fig. 2, right-hand column). See the key in Fig. 2 for meanings of symbols. **b**, The artefact frequency distribution is that determined from the Mungo B trench excavations¹⁷, and indicates maximum occupational density between 45 and 43 kyr. Mega-faunal extinction is shown to occur at about 46 kyr (refs 14, 15), which is shortly after initial human arrival at Lake Mungo and coincident with fluctuations between lake-full and regionally dry conditions. **c**, Low lake levels are identified by weak soils that contain clay and silty sand blown from the partly exposed lake floor. The three Lower Mungo gravel bands were formed during three high water stands, when the lake overflowed, after the sand barrier had emerged above lake level³. During Upper Mungo time, when the system was dry for long intervals (thereby permitting clay deflation from the lake floor), detailed lake oscillations are difficult to specify. A brief return to high lake levels is suggested by cessation of deflation at about 30 kyr. **d**, The record of desert dust accretion constructed from microscopic analysis of Wüstenquarz grains²⁴ (grains cm^{-2} counted in thin-section) shows several weak soil-forming episodes between 48 and 42 kyr with major accretion after 40 kyr. Grain counts are not significant beyond 150 grains cm^{-2} . While lake level changes reflect hydrologic controls in the southeastern highlands, the dust index is a response to erosion of the westerly dunefields. These two independent lines of evidence indicate a major climatic change of regional extent at 40 kyr.

decline in occupational density thereafter reflects the human response to increasing fluctuations in lake levels and aridity.

The initial lake filling occurred at about 60 kyr ago (Fig. 3c), broadly coincident with an ice advance in the highland catchments²⁵. High lake levels persisted, with minor and short-lived episodes of aridity, to about 40 kyr ago. Between 50 and 40 kyr ago, three phases of beach gravels alternate with weakly developed soils, in which aeolian silt and lake-floor clay components provide evidence of lake level changes of brief or intermittent duration. These changes coincide with the arrival of people at Lake Mungo and with the Australia-wide disappearance of megafauna by about 46 kyr ago^{14,15}.

A major hydrologic change occurs after the third gravel (deposited about 42 kyr ago), with a phase of prolonged silt-clay deposition. This event signals substantial changes in catchment conditions, with the end of dominantly high lake levels at about 40 kyr marking the boundary between the Lower and Upper Mungo units. The dust component, weakly preserved in soil breaks as early as 48 kyr ago, increases dramatically across this boundary, reflecting intensified aridity in upwind dunes of the continental interior.

The new chronology has important implications for the age of the Lake Mungo Geomagnetic Excursion¹⁶. Dated to between 30 and 39 kyr ago by radiocarbon (calibrated ages) and thermoluminescence^{16,26,27}, the stratigraphic level of the original site lies near, or at, the Lower to Upper Mungo boundary³, dated here to about 40 kyr ago. If, in the light of future work, this correlation holds true, then the Lake Mungo excursion and the Laschamp event, a feature dated recently to near 41 kyr ago and thought to have lasted approximately 1,500 years (ref. 28), may fall in the same time interval.

The importance of Lake Mungo to world archaeology and human evolution has recently been accentuated by claims for direct dating of the Mungo III skeleton to 62 ± 6 kyr ago⁸ and for the extraction of mitochondrial DNA from these remains⁴. The Mungo I and III burials were both inserted near the Lower to Upper Mungo stratigraphic boundary. Mungo III was placed in a grave 80–100 cm deep, and traces of pelletal clay and dark, reworked soil were found in the grave fill. The grave was dug into sands, dated here to 42 ± 3 kyr ago, underlain by a thin band of pelletal clay derived from the lake floor (Fig. 2). The overlying unit sealing the grave is dated to 38 ± 2 kyr ago. Therefore, Mungo III was buried 40 ± 2 kyr ago, close to the age estimates from two previous studies^{5,6}, but 20 kyr later than the most recent claims^{7,8}.

Reasons for the 20-kyr age discrepancy invite speculation. One factor may involve uncertainties in U-migration²⁹, which is enhanced in carbonate-rich sedimentary deposits such as those studied here. U-migration can considerably affect the accuracy of 'open system' ESR and U-series ages, such as those used to date the Mungo III skeletal remains^{7,8}. The previous two optical ages for the Mungo III burial relate to sediments collected in uncertain stratigraphic contexts more than 300 m away from the burial site⁸.

The Mungo I cremation is indistinguishable in age from the Mungo III burial. Our new age constraints for Mungo I are 15–20 kyr older than the previous radiocarbon-based estimates^{2,9,10}, a difference we attribute to contamination of the radiocarbon samples by younger carbon.

Our study shows that humans were present at Lake Mungo as early as 50–46 kyr ago. We find no evidence to support claims for human occupation or burials near 60 kyr ago. A most significant climatic change of the last 60 kyr occurred near 40 kyr ago, when early lake-shore *Homo sapiens*, with culturally advanced mortuary practices, were forced to adapt to increasing aridity. The Lake Mungo evidence presented here provides a new chronological and environmental context for examining the interaction between the first human arrivals, the last of the megafauna, and the Australian landscape under conditions of major climatic change. □

Received 19 September; accepted 10 December 2002; doi:10.1038/nature01383.

1. Bowler, J. M. & Thorne, A. G. in *The Origin of the Australians* (eds Kirk, R. L. & Thorne, A. G.) 127–138 (Australian Institute of Aboriginal Studies, Canberra, 1976).
2. Bowler, J. M., Jones, R., Allen, H. & Thorne, A. G. Pleistocene human remains from Australia: a living site and human cremation from Lake Mungo, western New South Wales. *World Archaeol.* **2**, 39–60 (1970).
3. Bowler, J. M. Willandra Lakes revisited: environmental framework for human occupation. *Archaeol. Oceania* **33**, 120–155 (1998).
4. Adcock, G. J. et al. Mitochondrial DNA sequences in ancient Australians: implications for human origins. *Proc. Natl Acad. Sci. USA* **98**, 537–542 (2001).
5. Oyston, B. Thermoluminescence age determinations for the Mungo III human burial. Lake Mungo, southeastern Australia. *Quat. Sci. Rev.* **15**, 739–749 (1996).
6. Bowler, J. M. & Price, D. M. Luminescence dates and stratigraphic analyses at Lake Mungo: review and new perspectives. *Archaeol. Oceania* **33**, 156–168 (1998).
7. Simpson, J. J. & Grün, R. Non-destructive gamma spectrometric U-series dating. *Quat. Sci. Rev.* **17**, 1009–1022 (1998).
8. Thorne, A. et al. Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *J. Hum. Evol.* **36**, 591–612 (1999).
9. Bowler, J. M., Thorne, A. G. & Polach, H. A. Pleistocene man in Australia: age and significance of the Mungo skeleton. *Nature* **240**, 48–50 (1972).
10. Gillespie, R. Burnt and unburnt carbon: dating charcoal and burnt bone from the Willandra Lakes, Australia. *Radiocarbon* **39**, 225–236 (1997).
11. Roberts, R. G., Jones, R. & Smith, M. A. Thermoluminescence dating of a 50,000 year-old human occupation site in northern Australia. *Nature* **345**, 153–156 (1990).
12. Roberts, R. G. et al. The human colonisation of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. *Quat. Sci. Rev.* **13**, 575–583 (1994).
13. Turney, C. S. M. et al. Early human occupation at Devil's Lair, southwestern Australia 50,000 years ago. *Quat. Res.* **55**, 3–13 (2001).
14. Miller, G. H. et al. Pleistocene extinction of *Genyornis newtonii*: human impact on Australian megafauna. *Science* **283**, 205–208 (1999).
15. Roberts, R. G. et al. New ages for the last Australian megafauna: continent-wide extinction about 46,000 years ago. *Science* **292**, 1888–1892 (2001).
16. Barbetti, M. & McElhinny, M. Evidence of a geomagnetic excursion 30,000 yr BP. *Nature* **239**, 327–330 (1972).
17. Shawcross, W. Archaeological excavations at Mungo. *Archaeol. Oceania* **33**, 183–200 (1998).
18. Bowler, J. M. & Magee, J. W. Redating Australia's oldest human remains: a sceptic's view. *J. Hum. Evol.* **38**, 719–726 (2000).
19. Gillespie, R. & Roberts, R. G. On the reliability of age estimates for human remains at Lake Mungo. *J. Hum. Evol.* **38**, 727–732 (2000).
20. Grün, R. et al. Age of the Lake Mungo 3 skeleton, reply to Bowler & Magee and to Gillespie & Roberts. *J. Hum. Evol.* **38**, 733–741 (2000).
21. Huntley, D. J., Godfrey-Smith, D. I. & Thewalt, M. L. W. Optical dating of sediments. *Nature* **313**, 105–107 (1985).
22. Aitken, M. J. *An Introduction to Optical Dating* (Oxford Univ. Press, Oxford, 1998).
23. Hope, J. H. Pleistocene mammal extinctions: the problem of Mungo and Menindee, New South Wales. *Alcheringa* **2**, 65–82 (1978).
24. Radczewski, O. E. in *Recent Marine Sediments: A Symposium* (ed. Trask, P. D.) 496–502 (Special Publication 4, Society for Economic Paleontologists & Mineralogists, Tulsa, Oklahoma, 1960).
25. Barrows, T. T., Stone, J. O., Fifield, L. K. & Cresswell, R. G. Late Pleistocene glaciation of the Kosciuszko Massif, Snowy Mountains, Australia. *Quat. Res.* **55**, 179–189 (2001).
26. Huxtable, J. & Aitken, M. J. Thermoluminescent dating of Lake Mungo geomagnetic polarity excursion. *Nature* **265**, 40–41 (1977).
27. Bell, W. T. Thermoluminescence dates for the Lake Mungo Aboriginal fireplaces and the implications for radiocarbon dating. *Archaeometry* **33**, 43–50 (1991).
28. Laj, C., Kissel, C., Mazaud, A., Channell, J. E. T. & Beer, J. North Atlantic palaeointensity stack since 75 ka (NAPIS-75) and the duration of the Laschamp event. *Phil. Trans. R. Soc. Lond. A* **358**, 1009–1025 (2000).
29. Ivanovich, M. & Harmon, R. S. (eds) *Uranium-series Disequilibrium: Applications to Earth, Marine, and Environmental Sciences*, 2nd edition (Clarendon, Oxford, 1992).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This project was funded by a grant from the Commonwealth Department of Environment and Heritage (administered by the NSW National Parks and Wildlife Service) with the support of the three traditional tribal groups of the Willandra Lakes World Heritage Area. Delayed neutron activation analyses were assisted by a grant from the Australian Institute of Nuclear Science and Engineering. We thank M. Pappin for co-ordinating indigenous contacts; M. Cupper, P. Hope, R. Gillespie, G. Robertson, J. Pappin, B. Mitchell, J. Kennedy and P. Lawson for field assistance; N. Hill, D. Questiaux and H. Yoshida for laboratory assistance. We also thank the Australian Research Council for supporting R.G.R. with a Senior Research Fellowship.

Author contributions J.M.B. managed the overall project and the geological components, and wrote the initial draft of this report; R.G.R. and N.A.S. carried out the OSL analyses; J.M.O., J.R.P. and N.A.S. made the dosimetry measurements; and H.J. and W.S. supervised the archaeological components.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.B. (e-mail: j.bowler@earthsci.unimelb.edu.au).

Architecture and material properties of diatom shells provide effective mechanical protection

Christian E. Hamm*, Rudolf Merkel†‡, Olaf Springer§, Piotr Jurkojc§, Christian Maier†, Kathrin Prechtel† & Victor Smetacek*

* Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

† Technische Universität München, Physics Department (Biophysics Group E22), 85748 Garching, Germany

‡ Research Centre Jülich, Institute of Thin Films and Interfaces, 52425 Jülich, Germany

§ Hochschule Bremen – University of Applied Science, Neustadtswall 30, 28199 Bremen, Germany

Diatoms are the major contributors to phytoplankton blooms in lakes and in the sea and hence are central in aquatic ecosystems and the global carbon cycle¹. All free-living diatoms differ from other phytoplankton groups in having silicified cell walls in the form of two ‘shells’ (the frustule) of manifold shape and intricate architecture² whose function and role, if any, in contributing to the evolutionary success of diatoms is under debate^{3–5}. We explored the defence potential of the frustules as armour against predators by measuring their strength. Real and virtual loading tests (using calibrated glass microneedles and finite element analysis) were performed on centric and pennate diatom cells. Here we show that the frustules are remarkably strong by virtue of their architecture and the material properties of the diatom silica. We conclude that diatom frustules have evolved as mechanical protection for the cells because exceptional force is required to break them. The evolutionary arms race between diatoms and their specialized predators will have had considerable influence in structuring pelagic food webs and biogeochemical cycles.

The striking shapes and complex patterns of the silica shells of recent and fossil diatoms^{2,6} bear resemblance to statically sophisticated (that is, stable and lightweight) man-made constructions⁷, suggesting that the frustules physically protect the cells against mechanical challenges. Because a large variety of organisms, ranging from parasitoid and ingesting protists to crustacean zooplankton, successfully feed on diatoms, the original inferred defensive function of the frustules⁸ faded into the background of diatom research. However, in an environment subject to heavy grazing pressure, any mechanism that reduces population mortality qualifies as an effective defence or deterrent. Thus, the course of phytoplankton evolution must be significantly shaped by the ‘arms race’, and external mechanical protection would be an effective first line of defence⁹.

To test this hypothesis we measured the forces necessary to break single, living cells of three representative, bloom-forming diatom species—the centric species *Thalassiosira punctigera*, the much larger *Coscinodiscus granii* and the pennate species *Fragilariopsis kerguelensis*. The centric genera are ubiquitous components of diatom blooms and their pillbox shape is widely represented. Owing to the peculiar mode of diatom cell division, the average size of a growing population successively decreases in diameter¹⁰, so we compared the strengths of large and small cells of *T. punctigera* as well. We chose *F. kerguelensis* because it is a prominent species in the Southern Ocean and the dominant component of the diatom ooze accumulating under the Antarctic Circumpolar Current (ACC), which has been attributed to the exceptionally robust nature of its frustules¹¹.

We used calibrated glass microneedles to load and break the frustules with defined forces (Fig. 1a–f). *T. punctigera* cells resisted forces of up to 260 μN and 180 μN for cells of 50 μm and 100 μm

diameter, respectively (Fig. 1g). Mechanical strength and size were not only inversely related within a single species but also between different species: *C. granii* cells (130 μm diameter) were crushed at lower forces (up to 90 μN) than *T. punctigera*. *F. kerguelensis* frustules (longest axis 30 μm , Fig. 1d, f) broke at 730 μN (Fig. 1g). Assuming that the glass needle pressed on an area of 100 μm^2 , the pressures resisted ranged between about 1 and 7 N mm^{-2} (equivalent to 100–700 tonnes m^{-2}). Much the same forces were required to break frustules of individual cells of the same species or size group tested, regardless of where they were applied.

A three-dimensional finite element model (FEM) of the *F. kerguelensis* frustule (Fig. 2a) was constructed to simulate the glass needle crush tests and to examine the values and distributions of stress within the frustule. Loading the FEM of diatom frustules of *F. kerguelensis* with external pressures along the girdle region (side), the valve (top and bottom) or across the girdle region created homogeneous, and thus statically favourable, stress distributions within the frustule (Fig. 2b). External pressures on several small patches, simulating a mandible bite, were also well absorbed (Fig. 2c, d): stress concentration was deflected by the transversal ribs (costae), which smoothly absorbed the stress from the fragile areas in between, in which most of the areolae (pores) are located. Removing the ribs and distributing a corresponding amount of material evenly throughout the whole structure increased stress and deformation values in any of the load cases by more than 70%. Thus less than 60% of the force required to break the frustules of *F. kerguelensis* would be necessary to destroy such a ribless frustule.

The highest stress values within the frustule before failure, calculated with the FEM of *F. kerguelensis*, and using the load cases described above (a force of 750 μN on the valve and on the girdle bands), were remarkably high: more than 540 N mm^{-2} were

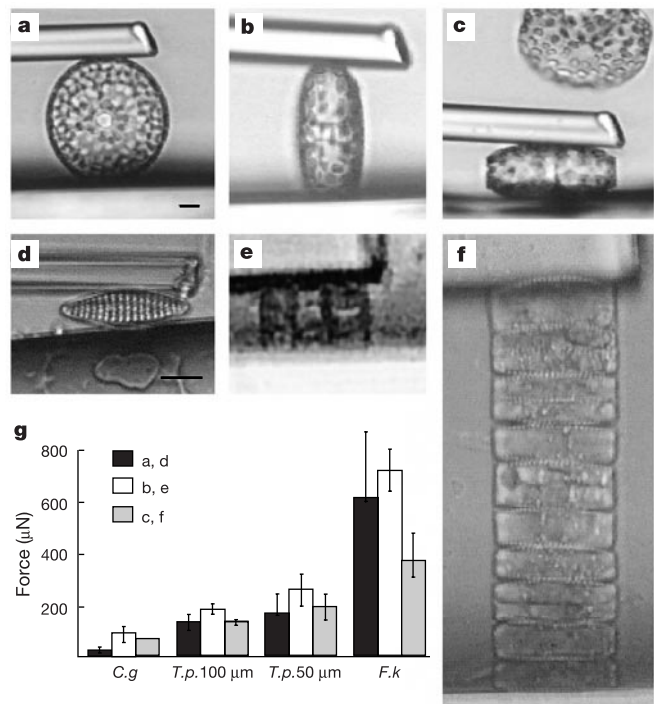


Figure 1 Glass needle tests: Live single cells of *T. punctigera* (a–c) and *F. kerguelensis* (d–f), in chains (e, f). Pressures applied along the girdle bands, (a, d), across the girdle bands (b, e), and across the centre of the valves (c, f). g, Forces necessary to break *Coscinodiscus granii* (*C.g.*), *Thalassiosira punctigera* (*T.p.*) with diameters of 100 and 50 μm , and *Fragilariopsis kerguelensis* (*F.k.*). *C. granii* has a geometry similar to that of *T. punctigera*. Scale bars, 10 μm .

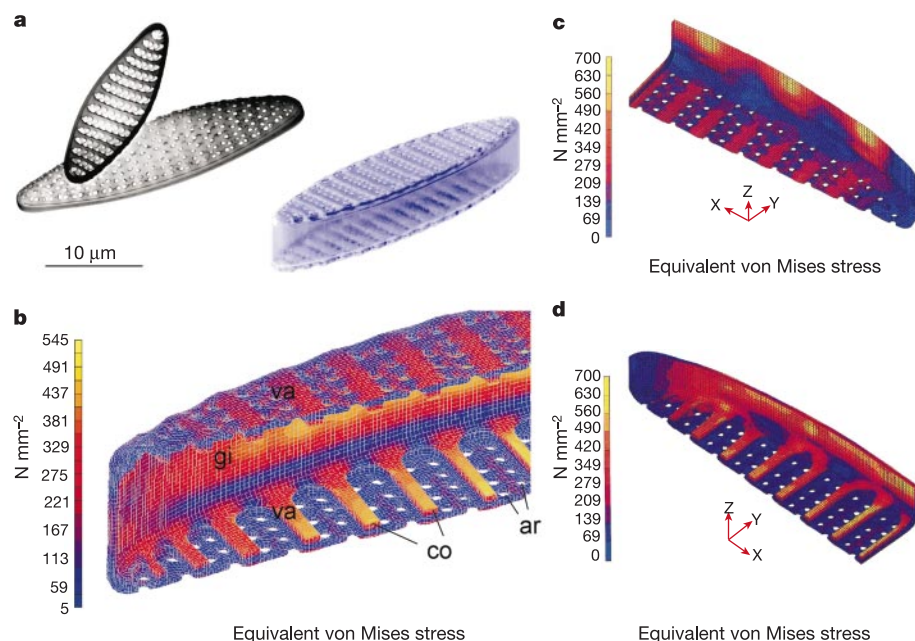


Figure 2 Finite element calculations of *F. kerguelensis* frustules. **a**, SEM of two valves (left) and rendered FEM of a complete frustule (right). **b**, Equivalent von Mises stress within the frustule as a function of pressure (total force, 750 μN) on the girdle region. gi, girdle region; va, valve; co, costae; ar, areolae. **c**, **d**, One-eighth sections, external (**c**) and

internal (**d**) views. Equivalent von Mises stress caused by pressure applied on several spots (area of each spot, 1 μm^2 ; total force, 750 μN), simulating a mandible bite; stress is absorbed by the costae.

reached before breakage (Fig. 2b). The values for tensile and compressive stress, depending on whether pressure was applied on the valve face or at the girdle region, were high within the girdle bands (155 and 330 N mm^{-2}) and in the costae (560 and 680 N mm^{-2}), suggesting that the silica forming the cell wall in this alga should have both high ultimate tensile and high ultimate compressive strengths.

Material properties of diatom silica were evaluated as a function of force by finite element simulation of the real deformation of an isolated pleura (a smooth, open, hoop-shaped segment of the girdle) of *T. punctigera* with a calibrated glass microneedle. The Young's modulus E of the diatom silica of a pleura was 22.4 GPa, which is comparable to cortical bone (20 GPa)¹² or medical dental composites (about 6–25 GPa)¹³. These consist, like diatom silica, of inorganic particles associated with an organic matrix^{14,15}. In

contrast, glass is much stiffer (about 70 GPa). Even very strong deformations of the pleura (2.5% strain, which causes, at $E = 22.4$ GPa, a stress of about 560 N mm^{-2}) elicited a completely elastic response. The original shape was regained when the pressure was removed. We were unable to break the pleurae with this method (Fig. 3).

The apparent absence of plastic deformation, which would indicate brittle failure, is consistent with the sharp-edged and smooth fracture sites that appear in broken frustules. However, a recent examination of a fracture in a pennate diatom has shown that the crack leading to fracture did not travel through, but around the minute (about 40 nm across) silica spheres forming the frustule¹⁶. This almost doubles the area of the fracture, thereby increasing the amount of energy necessary to break the frustule. In addition, several seconds elapsed before the frustules broke at a certain force,

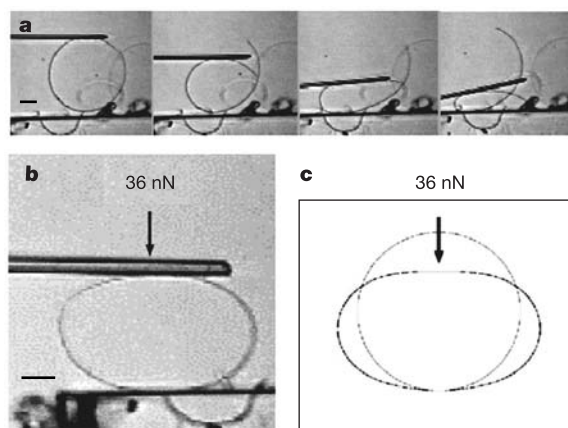


Figure 3 Properties of an isolated girdle band. **a**, Sequence showing strong elastic deformation of a girdle band as a function of increasing force. **b**, girdle band (pleura) deformed by a calibrated glass needle. **c**, FEM of the pleura, showing identical deflections

as functions of the same force (36 nN). The Young's modulus E of the silica in the FEM is 22.4 GPa. Scale bars, 10 μm .

indicating that creep occurred within the diatom silica, during which energy may have been dissipated by organic material within the composite, in analogy to a recently described process in bone¹⁷.

Clearly the frustules have evolved to withstand external pressure, and considerable energy and specialized tools are required to break them. Two major zooplankton groups in the ocean—copepods and euphausiids—have silica-edged mandibles and gizzards lined with formidable arrays of sharp ‘teeth’, respectively^{18,19}. These tools are likely to have co-evolved with their diatom prey. Indeed, diatom cells can survive gut passage if they escape being crushed²⁰. To generate the approximately 750 μN necessary to break *F. kerguelensis* frustules, the muscle strands operating crushing tools would have to have a diameter of more than 50 μm (based on a value of the strongest insect and vertebrate muscles²¹). Obviously, only large copepods and euphausiids can break such frustules. Small copepods have been shown to select other prey items²².

Although we tested only three species for strength, one for statics and two for material properties, we expect that a combination of balanced statics (lightweight architecture) and toughness of the material is generally present in diatom frustules. In the evolutionary arms race, reducing the numbers of potential predators will bring obvious advantages. It is now widely accepted that diatoms dominate phytoplankton blooms because their mortality rates are lower than those of other, smaller algae with similar growth rates¹. Our results suggest that the silica frustule is indeed an effective armour against many potential predators, although other specialized predators such as some ciliates²³ and dinoflagellates²⁴ have evolved mechanisms to digest diatom cell contents without breaking or opening the frustules.

Because diatoms play a major role in the ocean’s biological carbon pump and in food chains leading to fish, a better understanding of the ecological factors selecting for the species composition of their blooms is called for. We hypothesize that the perplexing variety in shapes and protuberances present in pelagic diatoms have largely evolved as responses to specific predator types ranging from puncturing and ingesting protists to various types of metazooplankton. Diatom frustules fascinate the human eye because of their arresting symmetry, but also because they duplicate, at minute scales, structures and patterns familiar to everyday experience. It seems that these microstructures are not merely ‘art forms of nature’²⁵, but have evolved under selection principles similar to those prevailing in the macroscopic world. □

Methods

Glass needle tests

Diatoms were cultured at 0 °C (*F. kerguelensis*) and 15 °C (*T. punctigera*, *C. granii*) under moderate light intensities. Microneedles were drawn from borosilicate glass with a micropipette puller. Microneedles were of cylindrical shape and showed no visual irregularities. These microtools were mounted in a micromanipulator. Experiments were performed in a measurement chamber filled with sea water on the stage of an inverted light microscope equipped with long-distance lenses and differential interference contrast optics. Micrographs were recorded with a charge-coupled device (CCD) camera, stored on videotape, and analysed using the image processing software NIH-Image. Glass needles were calibrated with atomic force microscope (AFM) cantilevers of known spring constants, which resulted in needle deflection as linear functions of applied forces. The pressure on the diatoms was gradually increased until the cells broke. Thus, by measuring the deflection of the needle in the relaxed state and directly before the cells broke, we could determine the maximal forces the cells were able to resist. Thicker and stiffer needles, which exceeded the range where the procedure described above was practicable, were calibrated by directly relating their deflections to forces that were measured on a microbalance. On the basis of the symmetries of the diatoms, we applied pressure with a glass microneedle from three different positions: (1) along the girdle bands, (2) across the girdle bands, and (3) across the centre of the valves.

Finite element modelling and calculations

An all-hexahedral finite element mesh of *F. kerguelensis* was generated with the finite element program ‘MSC Marc Mentat’ on the basis of micrographs (scanning electron and light microscopy, SEM and LM), and volume elements (hex 8) were used to construct the geometry. Hexahedral elements are generally preferred over tetrahedra in engineering²⁶. Resolution was around 350 elements μm^{-3} . We assumed symmetry in three planes. The load cases applied in the crash tests were simulated applying face pressure on the areas

where the glass needle was in contact with the diatom frustule. The complexity of diatom architecture and limitations in computing power forced us to simplify some parts of the structures (for example, in the girdle band region). The latter problem became less severe since the first finite element calculations on diatom frustules²⁷, but is not solved yet. For the same reasons, the raphe of *F. kerguelensis*, a slit extending over one edge of the valve, was not included in the model. We assumed statical symmetry of the valve, as the statical disturbance of the raphe is typically compensated for by structural reinforcements. Thus, our conclusions regarding the functional principles of the structure and the material properties should hardly be influenced by these reductions to the essential components.

Young’s modulus

Organic matter was removed from diatoms in a monospecific culture of *T. punctigera*. Purified siliceous components of the frustules consisted of the valves, the valvocopulae, the copulae and the pleurae; the geometrically simple pleurae were chosen to perform load tests with glass microneedles. The geometry of the pleura was investigated using SEM images of the cross-section of a broken pleura. Deformation as a function of applied force was calculated using glass needle tests (as above). The pleura was subsequently modelled using plate elements with a thickness of 140 and 70 nm. In accordance with the real load tests, the model was loaded with 10–40 nN within the finite element program under an assumed *E*-modulus. Deformation as a function of force was calculated and the *E*-modulus modified until real and virtual tests yielded identical results.

Received 29 August; accepted 24 December 2002; doi:10.1038/nature01416.

1. Smetacek, V. Diatoms and the ocean carbon cycle. *Protist* **150**, 25–32 (1999).
2. Round, F. E., Crawford, R. M. & Mann, D. G. *The Diatoms: Biology and Morphology of the Genera* (Cambridge Univ. Press, Cambridge, UK, 1990).
3. Tilman, D. & Kilham, P. Sinking in freshwater phytoplankton: Some ecological implications of cell nutrient status and physical mixing processes. *Limnol. Oceanogr.* **21**, 409–417 (1976).
4. Hale, M. S. & Mitchell, J. G. Functional morphology of diatom frustule microstructures: hydrodynamic control of Brownian particle diffusion and advection. *Aquat. Microb. Ecol.* **24**, 287–295 (2001).
5. Milligan, A. J. & Morel, F. M. M. A proton buffering role for silica in diatoms. *Science* **297**, 1848–1850 (2002).
6. Gersonde, R. & Harwood, D. M. Lower Cretaceous diatoms from ODP Leg 113 site 693 (Weddell Sea). Part 1: vegetative cells. *Proc. ODP Sci. Res.* **113**, 365–402 (1990).
7. Burkhardt, B. & Bach, K. *Diatoms 1—Shells in Nature and Techniques* (Cramer, Braunschweig, 1984).
8. Ehrenberg, C. G. *Die Infusionstierchen als vollkommene Organismen. Ein Blick in das tiefere organische Leben der Natur* (Leopold Voss, Leipzig, 1838).
9. Smetacek, V. A watery arms race. *Nature* **411**, 745 (2001).
10. Round, F. E. The problem of reduction of cell size during diatom cell division. *Nova Hedwigia* **23**, 291–303 (1972).
11. Verity, P. G. & Smetacek, V. Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Mar. Ecol. Prog. Ser.* **130**, 277–293 (1996).
12. Ashman, R. B., Cowin, S. C., Van Buskirk, W. C. & Rice, J. C. A continuous wave technique for the measurement of the elastic properties of cortical bone. *J. Biomech.* **17**, 349–361 (1984).
13. Willems, G., Lambrechts, P., Braem, M., Celis, J. P. & Vanherle, G. A classification of dental composites according to their morphological and mechanical characteristics. *Dent. Mater.* **8**, 310–319 (1992).
14. Kröger, N., Deutzmann, R. & Sumper, M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* **286**, 1129–1132 (1999).
15. Fortin, D. & Vargas, M. A. The spectrum of composites: new techniques and materials. *J. Am. Dental Assoc.* **131**, 26–30 (2000).
16. Crawford, S. A., Higgins, M. J., Mulvaney, P. & Wetherbee, R. Nanostructure of the diatom frustule as revealed by atomic force and scanning electron microscopy. *J. Phycol.* **37**, 543–554 (2001).
17. Thompson, J. B. et al. Bone indentation recovery time correlates with bond reforming time. *Nature* **414**, 773–776 (2001).
18. Sullivan, B. K., Miller, C. B., Peterson, W. T. & Soeldner, A. H. A scanning electron microscopy study of the mandibular morphology of boreal copepods. *Mar. Biol.* **30**, 175–182 (1975).
19. Suh, H. L. The gastric mill of euphausiid crustaceans: a comparison of eleven species. *Hydrobiologia* **321**, 235–244 (1996).
20. Fowler, S. W. & Fisher, N. S. Viability of marine phytoplankton in zooplankton fecal pellets. *Deep-Sea Res.* **30**, 963–969 (1983).
21. Weis-Fogh, T. & Alexander, R. M. in *Scale Effects in Animal Locomotion* (ed. Pedley, T. J.) 511–525 (Academic, London, 1977).
22. Atkinson, A. Omnivory and feeding selectivity in five copepod species during spring in the Bellingshausen Sea, Antarctica. *ICES J. Mar. Sci.* **52**, 385–396 (1995).
23. Smetacek, V. The annual cycle of protozooplankton in the Kiel Bight. *Mar. Biol.* **63**, 1–11 (1981).
24. Jacobson, D. M. A brief history of dinoflagellate feeding research. *J. Eukaryot. Microbiol.* **46**, 376–381 (1999).
25. Haeckel, E. *Kunstformen der Natur* (Bibliographisches Institut, Leipzig/Vienna, 1904).
26. Cifuentes, A. O. & Kalbag, A. A performance study of tetrahedral and hexahedral elements in 3-D finite element structural analysis. *Finite Elements Anal. Des.* **12**, 313–318 (1992).
27. Steucke, G. & Schmid, A.-M. M. in *Beiträge zum 1. internationalen Symposium des Sonderforschungsbereiches 230 Natürliche Konstruktionen, Leichtbau in Architektur und Natur Teil 2* 194–203 (Sprint-Druck GmbH, Stuttgart, 1989).

Acknowledgements We thank F. Hinz and R. Crawford for the SEM image of *F. kerguelensis*. U. Riebesell, D. Wolf-Gladrow, U. Bathmann and R. Crawford provided comments.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.E.H. (e-mail: chamm@awi-bremerhaven.de).

Brain-state- and cell-type-specific firing of hippocampal interneurons *in vivo*

Thomas Klausberger*, Peter J. Magill*, László F. Márton*, J. David B. Roberts*, Philip M. Cobden*, György Buzsáki† & Peter Somogyi*

*MRC Anatomical Neuropharmacology Unit, Department of Pharmacology, Oxford University, Mansfield Road, Oxford OX1 3TH, UK

†Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, New Jersey 07102, USA

Neural-network oscillations at distinct frequencies have been implicated in the encoding, consolidation and retrieval of information in the hippocampus. Some GABA (γ -aminobutyric acid)-

containing interneurons fire phase-locked to theta oscillations (4–8 Hz) or to sharp-wave-associated ripple oscillations (120–200 Hz), which represent different behavioural states^{1–6}. Interneurons also entrain pyramidal cells *in vitro*⁷. The large diversity of interneurons^{8–10} poses the question of whether they have specific roles in shaping distinct network activities *in vivo*. Here we report that three distinct interneuron types—basket, axo-axonic and oriens–lacunosum-moleculare cells—visualized and defined by synaptic connectivity as well as by neurochemical markers, contribute differentially to theta and ripple oscillations in anaesthetized rats. The firing patterns of individual cells of the same class are remarkably stereotyped and provide unique signatures for each class. We conclude that the diversity of interneurons, innervating distinct domains of pyramidal cells¹¹, emerged to coordinate the activity of pyramidal cells in a temporally distinct and brain-state-dependent manner.

Network activity patterns of the hippocampus include theta oscillations (4–8 Hz), which are observed in the rat during exploration and rapid-eye-movement sleep^{1,6}, and sharp-wave-associated

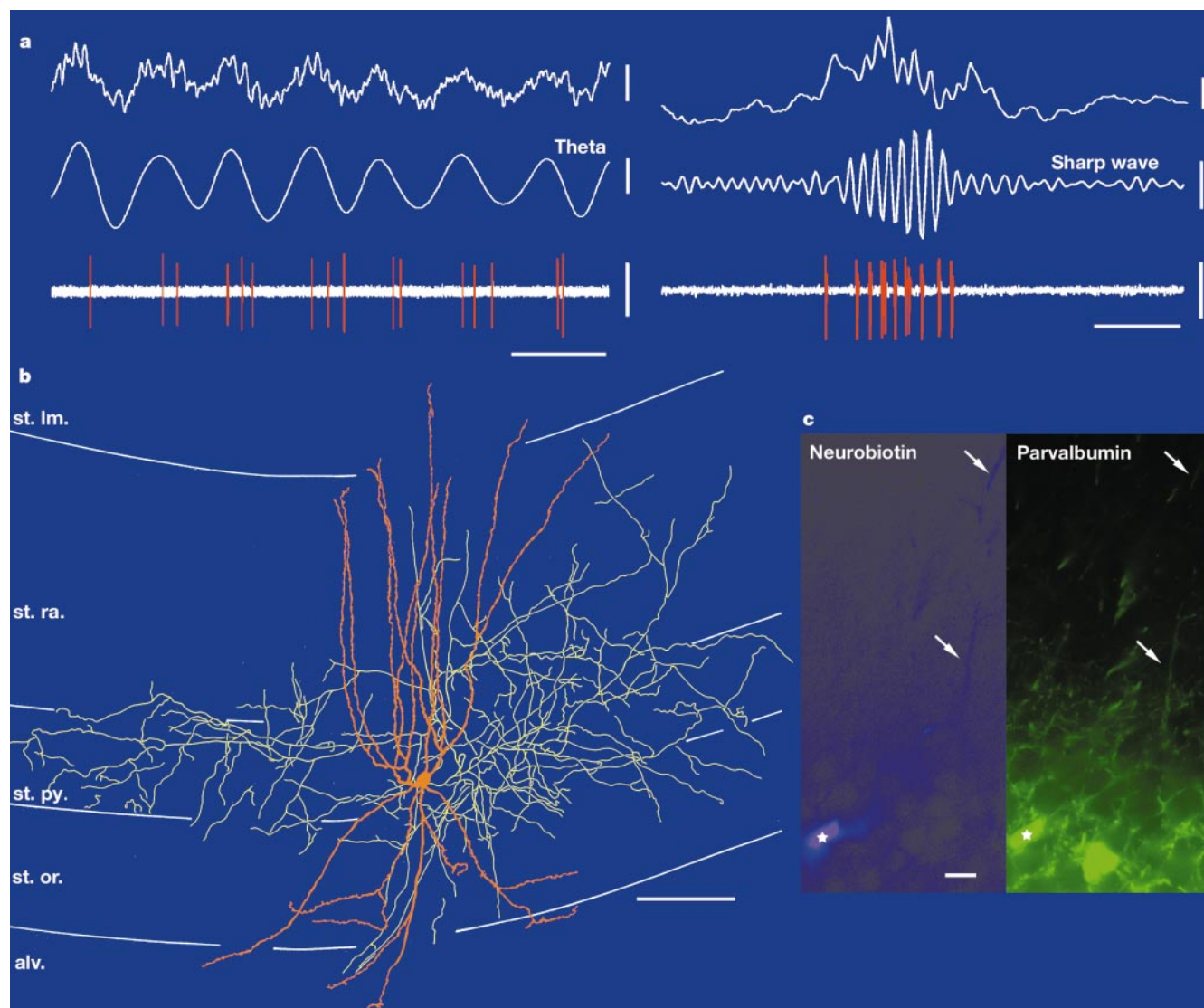


Figure 1 Firing patterns of a parvalbumin-positive basket cell (T44b) *in vivo*. **a**, The cell fired rhythmically on the descending phase of theta oscillations (filtered 3–6 Hz), and at high frequency at the troughs of sharp-wave-associated ripples (filtered 90–140 Hz). Scales: unit, 0.5 mV; unfiltered local field potential (LFP), 0.5 mV; theta, 0.3 mV, 300 ms; sharp wave, 0.2 mV, 50 ms. **b**, Reconstruction of the recorded and neurobiotin-labelled cell; soma and dendrites in red (complete). The axon in yellow, shown from only two

sections of 65 μ m thickness, contacted mainly pyramidal cell somata and proximal dendrites. **c**, The soma (star; partly seen) was cut during sectioning of the brain and is parvalbumin-immunopositive together with the dendrites (arrows) in immunofluorescence. alv., alveus; st. lm., stratum lacunosum-moleculare; st. or., stratum oriens; st. py., stratum pyramidale; st. ra., stratum radiatum. Scale bars: **b**, 100 μ m; **c**, 20 μ m.

ripples (120–200 Hz) that occur during slow-wave sleep, awake immobility and consummatory behaviours^{3,5,12}. We recorded the spiking activity of GABA-releasing interneurons from the dorsal CA1 region of the hippocampus in anaesthetized rats during network oscillations. Under the combined urethane–ketamine anaesthesia used here, high-frequency ripples (mean $\pm$ s.d., 123 ± 11 Hz) occurred spontaneously, and theta oscillations (4.2 ± 0.3 Hz) occurred after foot-pinch or spontaneously, depending on the depth of anaesthesia. Importantly, systemic administration of atropine (75 mg kg^{-1}) did not abolish theta oscillations (4.0 ± 0.3 Hz) in five rats tested. The presence of atropine-resistant theta is similar to that observed in the behaving rat and different from rats deeply anaesthetized with urethane¹³. The application of the juxtacellular method¹⁴ enabled us to label selectively the recorded neurons and to confirm their identity. Three distinct interneuron classes have been analysed by light and electron

microscopy, as well as by immunocytochemical testing for neuropeptides, calcium-binding proteins and neurotransmitter receptors.

Parvalbumin-expressing basket cells ($n = 5$) fired preferentially on the descending phase of the extracellular theta oscillations recorded in the stratum pyramidale. They fired often with one or several spikes (Fig. 1a) per ripple episode, at $12^\circ \pm 69^\circ$ (mean angle $\pm$ angular deviation) after the troughs of the ripple waves. The axon of basket cells was restricted mainly to the stratum pyramidale, or it also spread into neighbouring layers (Fig. 1b), but it predominantly contacted somata and proximal dendrites of pyramidal cells and other interneurons¹¹. The somata of two basket cells were located in the stratum pyramidale; the other three somata were in stratum oriens, including one next to the alveus. No physiological difference was observed between cells that had different soma locations. All basket cells were immunopositive for

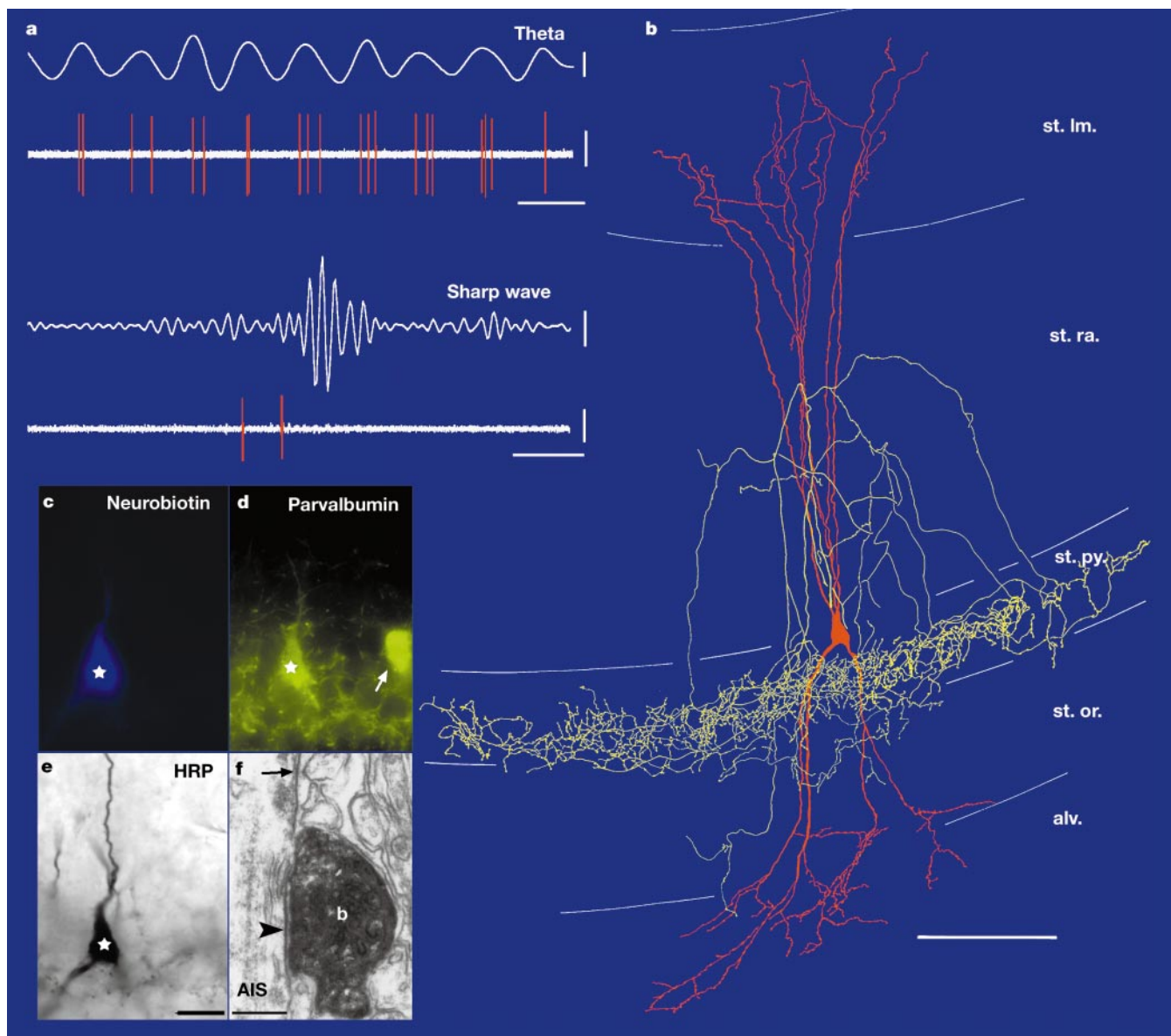


Figure 2 Firing patterns of a parvalbumin-positive axo-axonic cell (T76b) *in vivo*. **a**, The cell fired rhythmically just after the positive peak of theta oscillations, and at the beginning of a sharp-wave burst, but was silent subsequently. Scales: unit, 0.5 mV; theta, 0.2 mV, 300 ms; sharp wave, 0.1 mV, 50 ms. **b**, Reconstruction of the neurobiotin-labelled cell; soma and dendrites in red (complete); the axon (yellow) is restricted to the basal part of the stratum pyramidale and is shown from only one section of 65- μm thickness.

c, d, Immunofluorescence of parvalbumin in the axo-axonic (star) and in an unrecorded cell (arrow). **e**, Micrograph of the labelled cell. **f**, Electron micrograph showing a bouton (b) making a synapse (arrowhead) with an axon initial segment (AIS) characterized by membrane undercoating (small arrow). Abbreviations as in Fig. 1. Scale bars: **b**, 100 μm , **c–e**, 20 μm ; **f**, 0.2 μm .

parvalbumin (Fig. 1c), but were negative for the neuropeptide cholecystokinin (CCK).

Axo-axonic cells ($n = 2$) fired preferentially just after the peak of the theta cycles and discharged transiently at the beginning of sharp-wave-associated ripples (Fig. 2a). The somata of both cells were located in stratum pyramidale. In contrast to the basket cells, the dendrites of the axo-axonic cells branched extensively in stratum lacunosum-moleculare and in the alveus (Fig. 2b). Electron microscopic analysis confirmed the identity of these cells (Fig. 2f), as their postsynaptic targets were exclusively axon initial segments (13 synapses, cell T76b; eight synapses, cell T88a) of pyramidal cells¹¹. Both axo-axonic cells were strongly immunopositive for parvalbumin (Fig. 2d); one of the cells was tested for CCK and was immunonegative (data not shown).

Oriens-lacunosum-moleculare (O-LM) cells ($n = 3$) fired rhythmically at the trough of theta cycles and were suppressed during ripple episodes (Fig. 3a). The somata of all three cells were located in the stratum oriens; one cell was at the border of the CA1 area and subiculum (Fig. 3b). They had mainly horizontally running dendrites. The main axons of these cells crossed stratum pyramidale and radiatum, and branched heavily in stratum lacunosum-moleculare, where their terminals are co-aligned with the entorhinal input^{15–18}. All three O-LM cells were immunopositive for the metabotropic glutamate receptor mGluR1 α and the neuropeptide somatostatin, and they were also weakly positive for parvalbumin (Fig. 3c). In addition, one of the cells was tested with an antibody against mGluR7a, and the soma and dendrites were selectively decorated with strongly immunopositive terminals (Fig. 3d).

To compare the firing patterns of different neurons quantitatively⁵ (see also Supplementary Information), data from three different brain states were analysed: theta oscillation, sharp-wave/ripple epochs and non-theta/non-sharp-wave periods (Fig. 4a). The discharge frequency of pyramidal cells showed no significant difference (paired t -test, $P > 0.1$, $n = 6$) between theta and non-theta/non-sharp-wave periods. Interneurons, as a single group,

increased their discharge frequency during theta oscillations (paired t -test, $P < 0.01$, $n = 10$). During ripples, both pyramidal cells and basket cells increased their discharge frequency (paired t -test, $P < 0.05$ and $P < 0.01$, respectively) in comparison to non-theta/non-sharp-wave periods. Axo-axonic cells showed no significant difference in discharge frequency (paired t -test, $P > 0.1$), and O-LM cells decreased their firing rates during sharp waves (paired t -test, $P < 0.05$).

An even more robust difference between interneuron classes emerged from the analysis of temporal structure during network oscillations (Fig. 4b). During theta oscillations, pyramidal cells fired at $20^\circ \pm 65^\circ$, parvalbumin basket cells at $271^\circ \pm 68^\circ$, axo-axonic cells at $185^\circ \pm 55^\circ$, and O-LM cells at $19^\circ \pm 57^\circ$ (mean angle $\pm$ angular deviation). The phase preferences of the three interneuron classes during theta oscillations were all significantly different from each other (Watson–Williams test, $P < 0.01$). Notably, the interspike interval within theta cycles that had ≥ 2 spikes was different for the three cell types (Kruskal–Wallis test with post-hoc Dunn tests, $P < 0.01$). The interspike intervals were 45 ± 25 ms (mean $\pm$ s.d.), 20 ± 17 ms and 76 ± 19 ms for basket, axo-axonic and O-LM cells, respectively. This indicates that axo-axonic cells and basket cells discharge at gamma frequency within theta cycles, in contrast to O-LM cells. During sharp-wave-associated ripples, pyramidal cells and basket cells fired most frequently at the maximum amplitude of the ripple episode, resulting in a single, robust peak in the cross-correlogram (Fig. 4b). Surprisingly, both axo-axonic cells showed a different firing pattern; they increased their firing probability at the beginning of the sharp-wave episode, but were silent at the maximum amplitude and after the sharp wave. All O-LM cells were silent during sharp waves. The firing patterns of cells in the same class were remarkably similar, providing highly predictive signatures for each class. In addition, the firing patterns of the three classes of interneuron correspond to firing patterns of unidentified neurons in the behaving animal⁵. All parvalbumin-positive basket, axo-axonic and O-LM cells resemble ‘single peak’, ‘biphasic’ and

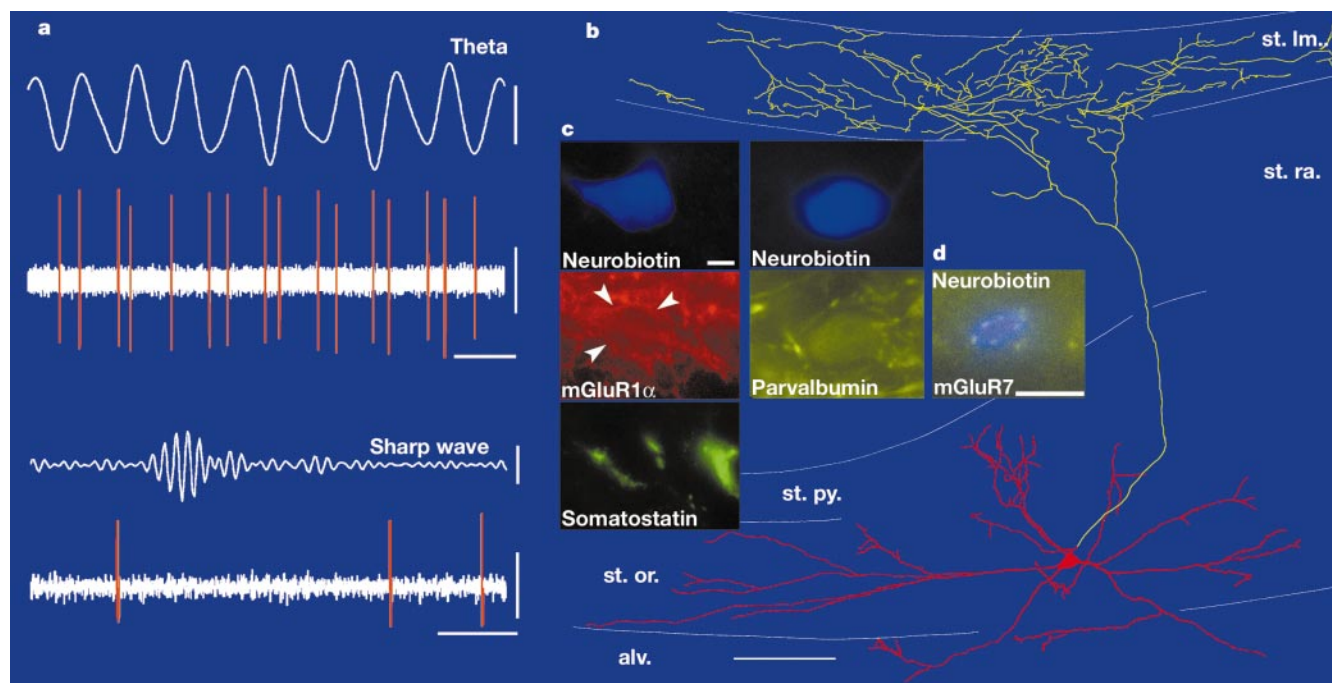


Figure 3 Firing patterns of an O-LM cell (T64a) *in vivo*. **a**, The cell fired rhythmically on the trough of theta oscillations, but was silent during sharp-wave-associated ripples. Scales: unit, 0.3 mV; theta, 0.2 mV, 300 ms; sharp wave, 0.05 mV, 50 ms. **b**, Reconstruction of the neurobiotin-filled oriens-lacunosum-moleculare (O-LM) cell. Soma and dendrites in red (complete); the axon restricted to str. lm. (yellow) is shown from only

two sections of 65- μ m thickness. **c**, The soma was cut into two sections; one shows somatostatin and metabotropic glutamate receptor mGluR1 α immunoreactivity; the other shows parvalbumin immunoreactivity. **d**, The dendrite of another recorded O-LM cell (T93b) is decorated with highly mGluR7a-immunopositive terminals (yellow). Abbreviations as in Fig. 1. Scales: **b**, 100 μ m; **c**, **d**, 5 μ m.

'anti-sharp-wave' neurons, respectively (see Methods). One of the O-LM cells did not reach significance for anti-sharp-wave neurons, because only a small number of ripples ($n = 10$) could be recorded. Nevertheless, the data confirmed that O-LM cells are silent during sharp waves.

We have shown that three distinct classes of interneuron, as

defined by their synaptic connectivity, contribute to different aspects of network oscillations *in vivo*. Therefore, it seems that a large diversity of GABA interneurons evolved to control pyramidal cells in a temporally distinct and brain-state-dependent manner. During theta activity, CA1 place cells¹ are activated primarily by the perforant path^{19,20}. The O-LM cells, whose axons are co-aligned with the perforant path input, fire coincidentally with the strongest hyperpolarization in the distal dendrites of pyramidal cells⁶. Therefore, O-LM cells might phase-modulate excitatory input from the perforant path²¹ and/or provide rhythmic hyperpolarization that could de-inactivate voltage-sensitive ion channels and facilitate somatodendritic back-propagation of the action potentials²² and burst discharge²³ in active place cells. Axo-axonic cells and basket cells have high discharge probabilities on the descending phase of the theta cycle at times when the discharge probability of pyramidal cells is lowest and gamma power is highest^{2,6,24–26}. Such cycle-phase arrangement can facilitate the rhythmic entrainment of pyramidal cell discharge⁷. During slow-wave sleep, the main driving source to CA1 pyramidal neurons is the sharp-wave-associated discharge of the CA3 pyramids⁵. Transient suppression of O-LM cell activity during sharp waves may allow action potentials in pyramidal cells to back-propagate to the most distal dendrites^{27,28} and facilitate long-term potentiation of those entorhinal input synapses that are coincidentally active during sharp waves. Basket cells (parvalbumin-positive) fire at high frequency and phase-locked to ripple oscillation, and can therefore provide an inhibitory temporal structure for large populations of pyramidal cells^{3,5}. By contrast, axo-axonic cells, which make inhibitory synapses at the site of action potential generation, only fire at the beginning of the sharp wave. Such a coordination of distinct inputs could time-lock the activity of thousands of pyramidal cells to fire together, exactly at the maximum amplitude of ripple episodes.

Different classes of interneuron innervate distinct domains of pyramidal cells and exhibit specific firing patterns during behaviourally relevant oscillations. Therefore, they probably evolved to temporally coordinate input–output transformation in the hippocampal pyramidal cells, and to govern the formation and retrieval of cell assemblies during hippocampus-dependent behaviour. □

Methods

Electrophysiological recordings

Rats were treated in accordance with the Animals (Scientific Procedures) Act, 1986 (UK) and associated procedures. Fifty-two male Sprague–Dawley rats (250–350 g) were anaesthetized with 1.25 mg kg^{−1} urethane, plus supplemental doses of ketamine and xylazine (20 and 2 mg kg^{−1}, respectively) as needed, and body temperature was retained with a heating pad. Neuronal activity in the hippocampus was recorded extracellularly with a glass electrode (18–25 MΩ) filled with 1.5% neurobiotin in 0.5 M NaCl, and the local field potential (LFP) was recorded with a stationary second electrode in the hippocampal CA1 pyramidal cell layer. Single-unit activity (sampling rate 20 kHz) and LFP (sampling rate 800 Hz) were filtered online at 0.8–5 kHz and 0.5–200 Hz, respectively. After the collection of a sufficient number of spikes, the electrode was advanced towards the cell, which was then labelled juxtacellularly with neurobiotin by applying positive current steps¹⁴. The shape and amplitude of spikes were monitored during recording, advancing the electrode and labelling to ensure that recorded spikes originated from a single neuron only, and that the recorded cell was labelled. Recordings from nine out of ten putative interneurons resulted in a single labelled interneuron following histological

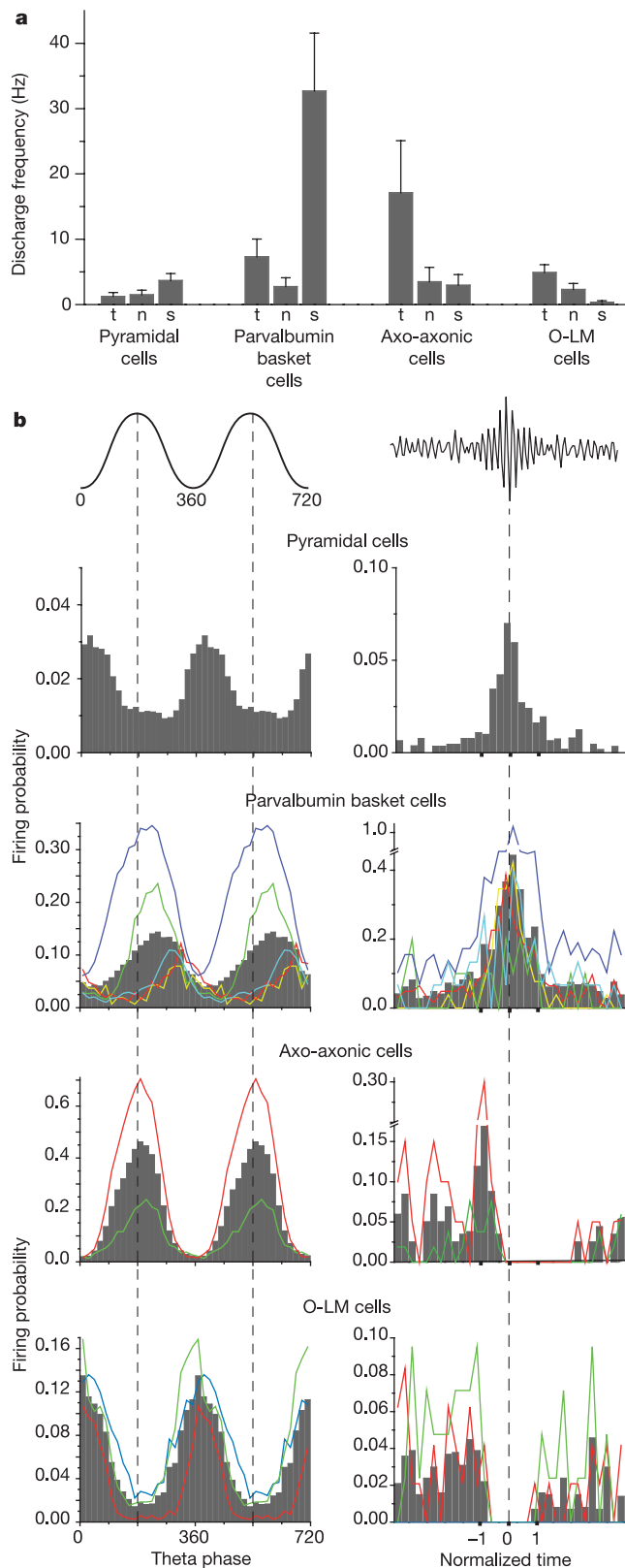


Figure 4 Distinct interneuron classes generate different firing patterns during theta and ripple oscillations *in vivo*. **a**, Discharge frequency (mean $\pm$ s.e.m.) of pyramidal cells, parvalbumin-positive basket cells, axo-axonic cells and O-LM cells during theta (t), non-theta/non-sharp-wave (n) and sharp-wave-associated ripples (s). **b**, The mean firing probabilities of different cell types are shown as grey columns, firing probabilities of each interneuron as coloured lines. For clarity, two theta cycles are shown; 0° and 360° mark the trough of the theta cycles recorded extracellularly in the stratum pyramidale (left column). The start, maximum amplitude and end of the normalized sharp-wave episodes are marked as −1, 0 and 1, respectively (right column). Note that the phase relationships of neurons to the network patterns are similar within the same class, despite variation in discharge probability of individual cells.

processing. In one case, spikes from a putative interneuron and a putative pyramidal cell were recorded simultaneously and, after labelling, a parvalbumin-positive basket cell and a pyramidal cell were recovered. The interneuron was more strongly modulated by the current steps and was also more strongly labelled than the pyramidal cell.

Anatomical and immunohistochemical visualization

The rats were perfused with fixative 4 h after labelling. Immunofluorescence and peroxidase reactions for light microscopy and electron microscopy were performed as described previously^{18,29}. Antibodies against parvalbumin, CCK, somatostatin and mGluR7a were gifts from K. Baimbridge, A. Varro, A. Buchan and R. Shigemoto, respectively; antibodies against mGluR1 were obtained from DiaSorin. The specificity of these antibodies is discussed elsewhere²⁹.

Data analysis

Theta epochs were detected by calculating the theta (3–6 Hz) to delta (2–3 Hz) frequency power ratio in 2-s windows of the LFP³. A ratio greater than four in at least three consecutive windows marked theta episodes, and a ratio less than two in at least three consecutive windows indicated epochs that are called non-theta/non-sharp-wave periods, which lacked field ripples. Non-theta/non-sharp-wave periods contained oscillations of 1 Hz and/or 2–3 Hz, but were not further analysed in this study, owing to the larger variability in cell activity. To determine the phase relationship between single-cell and theta activity, the troughs of the theta oscillations were detected in the filtered signal (3–6 Hz). Each spike was assigned to a given phase (bin size 20°) between the troughs (0° and 360°) and all theta cycles were superimposed⁵. Theta phase was analysed using circular statistics³⁰.

The LFP was filtered at 90–140 Hz for the detection of sharp-wave-associated ripples and the power (r.m.s. amplitude) of the filtered signal was calculated in 10-ms windows⁵. The threshold for ripple detection was set to 5 s.d. above the mean power. The beginning and the end of the sharp wave were set where the power crossed 1 s.d. above the mean power. The maximum amplitude of the oscillation was detected by a peak-finding algorithm. To evaluate the firing pattern of a single neuron during sharp waves, ripple episodes were normalized. Because sharp waves are often not symmetrical, the periods between the beginning and the ripple maximum, and the ripple maximum and the end of the sharp-wave episodes were each divided into four bins, and spikes were sorted into bins. To test whether potential correlations of firing patterns with sharp waves arose by chance, for each cell the exact number and duration of observed sharp-wave windows were randomly shuffled over the period of non-theta/non-sharp-wave epoch, and the spikes detected in the windows were binned. A neuron was considered to be a 'single peak' cell if the number of spikes in six bins surrounding the ripple maximum was higher in the observed sharp-wave correlogram than the mean + 2 s.d. from 100 shuffled correlograms. For anti-sharp-wave cells the number of spikes in six bins surrounding the ripple maximum was lower in the observed sharp wave correlogram than the mean – 2 s.d. from 100 shuffled correlograms. For biphasic neurons, the number of spikes in four bins surrounding the beginning of the sharp wave was higher in the observed sharp-wave correlogram than the mean + 1 s.d. from 100 shuffled correlograms, and the number of spikes in ten consecutive bins, starting from the fourth bin of the sharp wave, was lower in the observed sharp waves than the mean – 2 s.d. of the respective bins from 100 shuffled correlograms. The discharge frequency of single cells during three different brain states (theta, sharp-wave and non-theta/non-sharp-wave) was calculated by dividing the number of spikes with the summed duration of the respective brain state.

Received 15 October; accepted 29 November 2002; doi:10.1038/nature01374.

- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, UK, 1978).
- Fox, S. E. Membrane potential and impedance changes in hippocampal pyramidal cells during theta rhythm. *Exp. Brain Res.* **77**, 283–294 (1989).
- Ylinen, A. *et al.* Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: Network and intracellular mechanisms. *J. Neurosci.* **15**, 30–46 (1995).
- Ylinen, A. *et al.* Intracellular correlates of hippocampal theta rhythm in identified pyramidal cells, granule cells, and basket cells. *Hippocampus* **5**, 78–90 (1995).
- Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A. & Buzsáki, G. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J. Neurosci.* **19**, 274–287 (1999).
- Buzsáki, G. Theta oscillations in the hippocampus. *Neuron* **33**, 325–340 (2002).
- Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* **378**, 75–78 (1995).
- Freund, T. F. & Buzsáki, G. Interneurons of the hippocampus. *Hippocampus* **6**, 347–470 (1996).
- Fricker, D. & Miles, R. Interneurons, spike timing, and perception. *Neuron* **32**, 771–774 (2001).
- McBain, C. J. & Fisahn, A. Interneurons unbound. *Nature Rev. Neurosci.* **2**, 11–23 (2001).
- Buhl, E. H., Halasy, K. & Somogyi, P. Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites. *Nature* **368**, 823–828 (1994).
- Traub, R. D. *et al.* Axonal gap junctions between principal neurons: A novel source of network oscillations, and perhaps epileptogenesis. *Rev. Neurosci.* **13**, 1–30 (2002).
- Vanderwolf, C. H. Cerebral activity and behavior: Control by central cholinergic and serotonergic systems. *Int. Rev. Neurobiol.* **30**, 225–340 (1988).
- Pinaud, D. A novel single-cell staining procedure performed *in vivo* under electrophysiological control: Morpho-functional features of juxtacellularly labeled thalamic cells and other central neurons with biocytin or neurobiotin. *J. Neurosci. Methods* **65**, 113–136 (1996).
- McBain, C. J., DiChiara, T. J. & Kauer, J. A. Activation of metabotropic glutamate receptors differentially affects two classes of hippocampal interneurons and potentiates excitatory synaptic transmission. *J. Neurosci.* **14**, 4433–4445 (1994).
- Sik, A., Penttonen, M., Ylinen, A. & Buzsáki, G. Hippocampal CA1 interneurons: An *in vivo* intracellular labeling study. *J. Neurosci.* **15**, 6651–6665 (1995).
- Ali, A. B. & Thomson, A. M. Facilitating pyramid to horizontal oriens-alveus interneurone inputs: Dual intracellular recordings in slices of rat hippocampus. *J. Physiol. (Lond.)* **507**, 185–199 (1998).

- Maccaferri, G., Roberts, J. D. B., Szucs, P., Cottingham, C. A. & Somogyi, P. Cell surface domain specific postsynaptic currents evoked by identified GABAergic neurones in rat hippocampus *in vitro*. *J. Physiol. (Lond.)* **524**, 91–116 (2000).
- Nakazawa, K. *et al.* Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* **297**, 211–218 (2002).
- Brun, V. H. *et al.* Place cells and place recognition maintained by direct entorhinal–hippocampal circuitry. *Science* **296**, 2243–2246 (2002).
- Gillies, M. J. *et al.* A model of atropine-resistant theta oscillations in rat hippocampal area CA1. *J. Physiol. (Lond.)* **543**, 779–793 (2002).
- Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science* **268**, 297–300 (1995).
- Harris, K. D., Hirase, H., Leinekugel, X., Henze, D. A. & Buzsáki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141–149 (2001).
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
- Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746 (2002).
- Harris, K. D. *et al.* Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* **417**, 738–741 (2002).
- Katona, I., Acsády, L. & Freund, T. F. Postsynaptic targets of somatostatin-immunoreactive interneurons in the rat hippocampus. *Neuroscience* **88**, 37–55 (1999).
- Magee, J. C. & Johnston, D. A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209–213 (1997).
- Losonczy, A., Zhang, L., Shigemoto, R., Somogyi, P. & Nusser, Z. Cell type dependence and variability in the short-term plasticity of EPSCs in identified mouse hippocampal interneurons. *J. Physiol. (Lond.)* **542**, 193–210 (2002).
- Zar, J. H. *Biostatistical Analysis* (Prentice Hall, New Jersey, 1999).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank G. Horseman and S. Gray from Cambridge Electronic Design, and P. Jays and L. Norman for technical assistance. We thank Z. Nusser, G. Tamas and J. Csicsvari for critically reading an earlier version of the manuscript, and Y. Dalezios for help with the statistics. T.K. was supported by an Erwin Schroedinger Fellowship from the Austrian Science Fund during part of this study; G.B. was supported by the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.K.
(e-mail: thomas.klausberger@pharm.ox.ac.uk).

Yeast genome duplication was followed by asynchronous differentiation of duplicated genes

Rikke B. Langkjær*, Paul F. Cliften†, Mark Johnston† & Jure Piškur*

* BioCentrum-DTU, Technical University of Denmark, Building 301, DK-2800 Lyngby, Denmark

† Department of Genetics and Genome Sequencing Center, Washington University School of Medicine, St. Louis, Missouri 63110, USA

Gene redundancy has been observed in yeast, plant and human genomes, and is thought to be a consequence of whole-genome duplications^{1–3}. Baker's yeast, *Saccharomyces cerevisiae*, contains several hundred duplicated genes¹. Duplication(s) could have occurred before or after a given speciation. To understand the evolution of the yeast genome, we analysed orthologues of some of these genes in several related yeast species. On the basis of the inferred phylogeny of each set of genes, we were able to deduce whether the gene duplicated and/or specialized before or after the divergence of two yeast lineages. Here we show that the gene duplications might have occurred as a single event, and that it probably took place before the *Saccharomyces* and *Kluyveromyces* lineages diverged from each other. Further evolution of each duplicated gene pair—such as specialization or differentiation of

the two copies, or deletion of a single copy—has taken place independently throughout the evolution of these species.

Most genomes show a high degree of redundancy^{4–6}, which can arise from single-gene duplications, duplications of short chromosomal segments or of entire chromosomes, or by duplication of the entire genome. All of these events are believed to play an important part in biological evolution⁷. Two genes derived from a gene duplication are said to be paralogous. One of the paralogues is often subsequently deleted from the genome. Rarely, paralogues are preserved because they differentiate and become functionally specialized. In the fruitfly, the nematode worm and the fission yeast genomes, most paralogous genes are dispersed. In contrast, duplicated genes in the yeast *S. cerevisiae*¹, the plant *Arabidopsis thaliana*², and in humans, *Homo sapiens*³, often occur in large segmental duplications in which members of homologous gene pairs are located in the same order along the two distinct segments and are sporadically interspersed with unique genes.

Owing to its small size, high gene density and a paucity of introns⁸, the *S. cerevisiae* genome is ideal for studying the origins of genome redundancy. Since their discovery^{9,10}, the origin of duplicated blocks of genes in the yeast genome has been controversial. A single genome duplication followed by partial deletions has been proposed as the origin of the duplicated *S. cerevisiae* regions/blocks¹, which consist of 655 duplicated genes¹¹. However, a series of continuous, smaller duplications¹², or a combination of both types of event^{13,14} could result in a similar genome configuration. A recent detailed analysis of the chromosomal gene order in several partially sequenced ascomycetous yeasts¹⁵ showed that three-quarters of the *S. cerevisiae* genome consists of 'sister regions' originating from the ancient duplicated segments¹⁶. The conserva-

tion of transcriptional orientation of duplicated gene pairs with respect to the centromere, the lack of triplicated regions, and pairing of the *S. cerevisiae* centromeres, strongly support the hypothesis of the whole-genome duplication^{1,16}. However, the timing of the whole-genome duplication and the subsequent fate of duplicated genes remain unclear.

We determined, and analysed, the sequence of genes orthologous to some duplicated *S. cerevisiae* genes from several yeast species. These species were selected on the basis of their phylogenetic relationship with *S. cerevisiae* (ref. 17; Fig. 1a). *Saccharomyces bayanus*, a *sensu stricto* species, is one of the closest relatives of *S. cerevisiae*¹⁷. *Saccharomyces castellii*, a *sensu lato* species, branched out before the *S. cerevisiae* and *S. bayanus* lineages separated. These three yeasts can generate respiratory deficient mutants (petite-positive phenotype), can grow anaerobically, and in general possess very similar physiological properties¹⁸. In contrast, *Saccharomyces kluyveri* and *Kluyveromyces lactis* are petite-negative, and have a somewhat different metabolism; *S. kluyveri* can also grow anaerobically, but exhibits weaker glucose repression of the respiratory pathway^{19,20}; *K. lactis* is the least related¹⁷ to *S. cerevisiae*, and can only grow aerobically²¹. *Candida albicans*, the fifth species used in our analysis, is not a very close relative of the other species we analysed¹⁷. Thirty-eight genes, covering almost half of the 52 duplicated blocks in the *S. cerevisiae* genome¹¹ and representing one-tenth of the total number of the *S. cerevisiae* duplicated genes, were analysed for their phylogenetic relationship (Table 1, Fig. 1b). We assumed that only a negligible horizontal transfer of the genetic material has occurred after separation of these yeasts, because, in general, the gene phylogenetic trees recapitulated the species phylogeny.

When a gene from one of the yeast species was compared with the pair of duplicated *S. cerevisiae* genes, two different types of phylogenetic tree could be obtained, depending on whether the *S. cerevisiae* gene pair was duplicated and/or differentiated before or after the species carrying the gene in question separated from the *S. cerevisiae* lineage (Fig. 1b). If a gene grouped with one of the *S. cerevisiae* paralogues (tree type A; Fig. 1b), we inferred that the duplication or gene specialization event in *S. cerevisiae* took place before the separation of the yeast lineages, and therefore this gene is of type A. If a gene was positioned outside the duplicated *S. cerevisiae* gene pair (tree type B; Fig. 1b), we inferred that speciation of the two yeast lineages took place before the gene duplication or specialization event. Surprisingly, all species contained a mixture of A- and B-type genes (Table 1). The proportion of A-type genes decreased as the genetic distance between *S. cerevisiae* and the analysed yeast species became greater, being 95% for *S. bayanus*, 81% for *S. castellii*, 58% for *S. kluyveri*, 28% for *K. lactis* and 3% for *C. albicans*. Apparently, the organization of the duplicated blocks and genes is very similar between *S. cerevisiae* and *S. bayanus*, whereas *C. albicans* seems to be a pre-duplication lineage. The branch-position of a majority of *C. albicans* gene sequences agrees with the phylogenetic relationship among studied yeast species. However, in a few cases (for example, the *GUP* gene) the position is embedded within the tree, suggesting a limited horizontal transfer.

In general, the switch from B type to A type for a certain gene is highly ordered, and the gene status in each species is a function of the phylogenetic relationship among the studied species. In other words, all A-type genes found in *K. lactis* are also of A type in the other yeasts, all of which are more closely related to *S. cerevisiae*; all *S. kluyveri* A-type genes are also of A type in *S. bayanus* and *S. castellii*; and the same rule applies for all *S. castellii* A-type genes. Once a yeast lineage acquired a gene of A type, the nature of this gene was preserved in all descending lineages. In addition, among the duplication blocks, which were represented by more than one gene (block 2, 8, 11, 23, 30, 34, 37, 43, 46 and 54), roughly 40% consisted of an A-type gene and a B-type gene in at least one species. For example, block 23 was represented by two gene pairs,

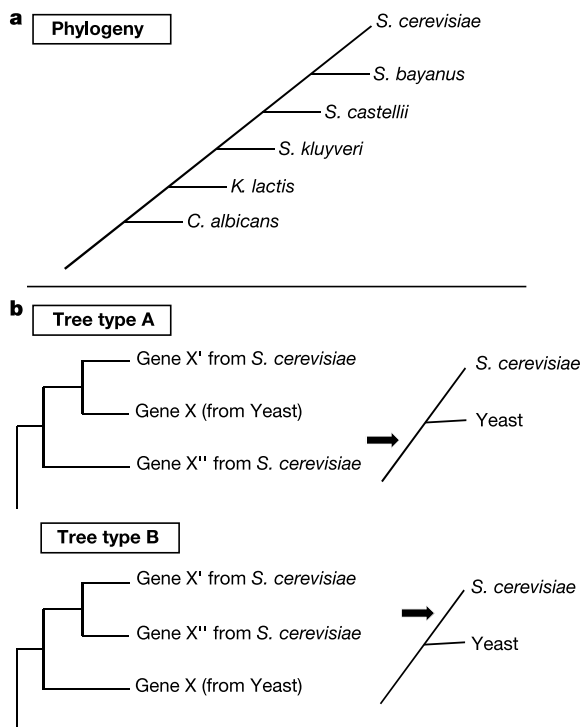


Figure 1 Phylogenetic relationship among different yeasts and among genes homologous to the duplicated *S. cerevisiae* genes. **a**, A schematic phylogenetic tree adapted from ref. 17. **b**, A yeast gene, homologous to the *S. cerevisiae* duplicated gene pair, can either be grouped with one of the duplicated genes (type A), or be outside the duplicated gene pair (type B). In the first case, the duplication or differentiation/specialization took place (see arrow) before the *S. cerevisiae* lineages and the studied yeast separated. In the case of tree type B, the duplication or differentiation/specialization event took place (see arrow) after separation of the lineages.

Table 1 Phylogenetic analysis of genes, homologous to selected *S. cerevisiae* paralogues

<i>S. cerevisiae</i>					Tree type					
Block	Gene	Systematic		Gene	Systematic	<i>S. b.</i>	<i>S. ca.</i>	<i>S. k.</i>	<i>K. l.</i>	<i>C. a.</i>
2	<i>MYO4</i>	YAL029C	-	<i>MYO2</i>	YOR326W	A	A	A	-	n.r.
2	<i>CDC19/PYK1</i>	YAL038W	-	<i>PYK2</i>	YOR347C	A	A	A	A	n.r.
3	<i>BAP2</i>	YBR068C	-	<i>BAP3</i>	YDR046C	A	A	B	-	-
8	<i>SMY2</i>	YBR172C	-	<i>YPL105C</i>	YPL105C	A	A	A	B	B
8	<i>SSE2</i>	YBR169C	-	<i>SSE1</i>	YPL106C	A	A	A	-	B
10	<i>PHO87</i>	YCR037C	-	<i>PHO90</i>	YJL198W	A	B	B	-	B
11	<i>CIT2</i>	YCR005C	-	<i>CIT1</i>	YNR001C	A	A	A	A	B*
11	<i>ADY2</i>	YCR010C	-	<i>FUN34</i>	YNR002C	A	A	B	-	B
11	<i>ARE1</i>	YCR048W	-	<i>ARE2</i>	YNR019W	A	A	A	-	B
11	<i>SOL2</i>	YCR073WA	-	<i>SOL1</i>	YNR034W	A	A	A*	-	B
21	<i>SIR2</i>	YDL042C	-	<i>HST1</i>	YOL068C	A	A*	B	B	B
22	<i>AKR1</i>	YDR264C	-	<i>AKR2</i>	YOR034C	A	A	A	A	B
23	<i>EFT2</i>	YDR385W	-	<i>EFT1</i>	YOR133W	B	B	B	-	B
23	<i>PDR15</i>	YDR406W	-	<i>PDR5</i>	YOR153W	A	A	n.r.	B	B
27	<i>RPS24A</i>	YER074W	-	<i>RPS24B</i>	YIL069C	B	B	B	B	B
28	<i>GEA2</i>	YEL022W	-	<i>GEA1</i>	YJR031C	A	A	-	-	B
29	<i>GND2</i>	YGR256W	-	<i>GND1</i>	YHR183W	A	A	-	B	B
30	<i>GSC2/GLS2</i>	YGR032W	-	<i>FKS1/GLS1</i>	YLR342W	A	B*	B*	-	B
30	<i>YGR043C</i>	YGR043C	-	<i>TAL1</i>	YLR354C	A	A	A*	A*	B
33	<i>GUP1</i>	YGL084C	-	<i>GUP2</i>	YPL189W	A	A	A	A	A
34	<i>DBF2</i>	YGR092W	-	<i>DBF20</i>	YPR111W	A	n.r.	B	B	B
34	<i>RPS23A</i>	YGR118W	-	<i>RPS23B</i>	YPR132W	A	B*	B	B	B
37	<i>TOM71/TOM72</i>	YHR117W	-	<i>TOM70</i>	YNL121C	A	A	A	-	B
37	<i>YCK1</i>	YHR135C	-	<i>YCK2</i>	YNL154C	A	A	-	B	B
38	<i>UBP7</i>	YIL156W	-	<i>UBP11</i>	YKR098C	A	A	A	-	B
39	<i>SEC24</i>	YIL109C	-	<i>SFB2</i>	YNL049C	A	A	A	-	B
42	<i>YJR061W</i>	YJR061W	-	<i>MNN4</i>	YKL201C	A	A	B	B	B
43	<i>GFA1</i>	YKL104C	-	<i>YMR085W</i>	YMR085W	A	A	A	-	n.r.
				<i>YMR084W</i>	YMR084W					
43	<i>YKL133C</i>	YKL133C	-	<i>YMR115W</i>	YMR115W	A	A	A	-	B*
45	<i>EXG1</i>	YLR300W	-	<i>SPR1</i>	YOR190W	A	A	A	B	B
46	<i>BUL2</i>	YML111W	-	<i>BUL1</i>	YMR275C	A	A	B	-	B
46	<i>ATR1</i>	YML116W	-	<i>YMR279C</i>	YMR279C	A	A*	A*	-	B
49	<i>CLA4</i>	YNL298W	-	<i>SKM1</i>	YOL113W	A	A	-	-	B
50	<i>YNL108C</i>	YNL108C	-	<i>TFC7</i>	YOR110W	A	A	A	-	B
54	<i>RPL13A</i>	YDL082W	-	<i>RPL13B</i>	YMR142C	A	B	B	B	B
54	<i>RPS16B</i>	YDL083C	-	<i>RPS16A</i>	YMR143W	A	B	B	B	B
Possible	<i>PET9/AAC2</i>	YBL030C	-	<i>AAC3</i>	YBR085W	A	A*	B*	B*	B*
Possible	<i>STP1</i>	YDR463W	-	<i>STP2</i>	YHR006W	A	A	A	-	-

The block numbers, the genes contained within them and the chromosome location can be found at <http://wolfe.gen.tcd.ie> and in ref. 11. For definition of tree type, see Fig. 1b. Further details of the analysis can be found in Methods; further details of the employed sequences, including the accession numbers, and the phylogenetic trees are available as Supplementary Information. Genes, where relevant branches were supported by bootstrap values of at least 750 out of 1,000, were either designated as A or B type, or in the case that the relevant branch was supported by values of 500–749 out of 1,000, as A* or B* type. *S. b.*, *S. bayanus*; *S. ca.*, *S. castellii*; *S. k.*, *S. kluyveri*; *K. l.*, *K. lactis*; *C. a.*, *C. albicans*; n.r., not resolved; -, sequences not available.

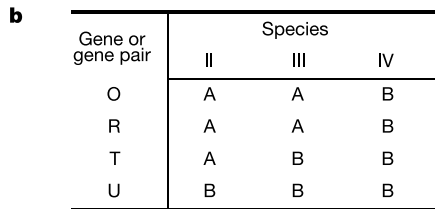
EFT1/EFT2 and *PDR15/PDR5*, but these two genes gave different tree types in *S. bayanus* and *S. castellii* (Table 1). Therefore, the fate of the duplicated genes has apparently not been the same, and there must have been an element of asynchrony during the evolutionary history of the yeast genome.

The asynchrony could be a consequence of serial duplications that took place continuously over a long period of time¹². However, the organization and relative location of all the duplicated blocks in *S. cerevisiae*¹ and the chromosomal gene order in several related species¹⁶ argue against this proposal, although the occurrence of small serial duplications and deletions is likely to have a minor role during the evolution of the *Saccharomyces* yeasts^{12,13,22,23}. Another explanation for the phylogenetic asynchrony observed among the analysed genes is that different gene pairs differentiated and specialized at different time points after the genome duplication.

In our present model (Fig. 2), we propose that a whole-genome duplication took place in the progenitor of the modern *Saccharomyces* and *Kluyveromyces* yeast lineages. Assuming a constant molecular clock for the 18S ribosomal RNA upon divergence of the Ascomycota and Basidiomycota about 550 million years ago²⁴, the duplication event is estimated to have taken place about 150 million years ago. This event provided the basis for speciation of different yeast lineages and for development of new metabolic life styles, including the ability of facultative fermentation, anaerobic growth, and repression by glucose¹⁴. However, during the era that passed until both copies of any pair had evolved enough to become indispensable for the genome, most paralogous gene pairs lost one of the copies. In addition, separation of different lineages often

occurred faster than the gene-pair differentiation/specialization process (Fig. 2). We need to take into account that the acquisition of a new function of a duplicated gene boosted accumulation of mutations, especially non-synonymous ones. We imagine that the predominantly A-type genes in Table 1 are those that specialized very soon after the duplication (for example, *CIT1/2*, *PYK1/2*, *AKR1/2*, *TAL1/YGR043C* and *GUP1/2*; Table 1, Fig. 2). Predominantly B-type genes would be those that retained the same function as their duplicated partner for a longer period, and which probably regularly exchanged sequence through gene conversion (for example, *PHO87/90*, *RPS23A/B*, *BUL1/2* and *RPL13A/B*) and did not specialize until a majority of lineages had speciated. The presence of both A- and B-type genes within the same block is also a relic of the asynchronous gene specialization process.

We note that our model (Fig. 2) does not contradict the observations used as the basis for the previously presented models of single-genome duplication¹⁶ and for a series of continuous duplications¹². The ancient genome duplication, asynchronous differentiation and deletion of genes have been accompanied by several other events, like extensive micro- and macro-reorganization of the chromosomes¹³ and minor duplication events^{22,23}, which have reshaped the yeast genomes into the contemporary forms. Initially, the number of centromeres was doubled, but only the *S. cerevisiae* and *S. bayanus* lineage retained the double chromosome number; other lineages apparently lost some of the duplicated centromeres during extensive chromosome reorganizations. A whole-genome duplication followed by asynchronous differentiation/deletion events, like that proposed here, might also explain



the reshaping of the genome and chromosomes of other species, including humans. □

Homologues/orthologues of the duplicated genes

NATURE | VOL 421 | 20 FEBRUARY 2003 | www.nature.com/nature

ublishing Group

drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci.* **10**, 569–570 (1994).
 30. Langkjaer, R. B. *Re-modelling of the Nuclear and Mitochondrial Genomes during the Evolution of the Saccharomyces Yeasts*. Thesis, BioCentrum-DTU, Technical Univ., Denmark (2002).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank the laboratories of C. Gaillardin and M. Bolotin-Fukuhara for providing *E. coli* clones containing genomic yeast DNA, which were partially sequenced by the Génolevures project. We thank K. H. Wolfe and his co-workers for sharing their manuscripts before publication. Sequencing of *Candida albicans* was accomplished with the support of the NIDR and the Burroughs Wellcome Fund. This work was partially supported by the Danish Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.P. (e-mail: jp@biocentrum.dtu.dk). The analysed gene sequences have the following GenBank accession numbers: AY144796–AY145050.

CD4⁺ T cells are required for secondary expansion and memory in CD8⁺ T lymphocytes

Edith M. Janssen^{*}, Edward E. Lemmens^{*}, Tom Wolfe[†], Urs Christen[†], Matthias G. von Herrath[†] & Stephen P. Schoenberger^{*}

^{*} Division of Cellular Immunology and [†] Division of Developmental Immunology, La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, California 92121, USA

A long-standing paradox in cellular immunology concerns the conditional requirement for CD4⁺ T-helper (T_H) cells in the priming of cytotoxic CD8⁺ T lymphocyte (CTL) responses *in vivo*. Whereas CTL responses against certain viruses can be primed in the absence of CD4⁺ T cells, others, such as those mediated through 'cross-priming' by host antigen-presenting cells, are dependent on T_H cells^{1–4}. A clearer understanding of the contribution of T_H cells to CTL development has been hampered by the fact that most T_H-independent responses have been demonstrated *ex vivo* as primary cytotoxic effectors, whereas T_H-dependent responses generally require secondary *in vitro* re-stimulation for their detection. Here, we have monitored the primary and secondary responses of T_H-dependent and T_H-independent CTLs and find in both cases that CD4⁺ T cells are dispensable for primary expansion of CD8⁺ T cells and their differentiation into cytotoxic effectors. However, secondary CTL expansion (that is, a secondary response upon re-encounter with antigen) is wholly dependent on the presence of T_H cells during, but not after, priming. Our results demonstrate that T-cell help is 'programmed' into CD8⁺ T cells during priming, conferring on these cells a hallmark of immune response memory: the capacity for functional expansion on re-encounter with antigen.

After exposure to antigen *in vivo*, CD8⁺ T-cell responses proceed through an ordered sequence of developmental events. These include an initial expansion phase in which a significant number of cytotoxic effectors are generated, and a subsequent contraction phase in which about 90% of these die, leaving a stable population of memory cells able to mount a rapid secondary response to antigen⁵. Although transition through these stages was thought to be governed by the presence and eventual elimination of antigen, mounting evidence suggests that CD8⁺ T-cell development is

guided by an instructional programme that, once set into motion during priming, is executed independently^{6–9}. The role of T_H cells in initiating and promoting the programmed development of CD8⁺ T cells *in vivo* is poorly understood. Early *in vitro* studies gave rise to models in which help is mediated by paracrine cytokines produced by T_H cells¹⁰. Recent work, however, has focused on the role of CD4⁺ T cells in activating antigen-presenting cells (APCs) from an immature state in which they are unable to prime CTLs to a state in which they can do so autonomously^{3,11,12}. T_H-independent antigens such as viruses are thought to prime CTLs by achieving a functionally equivalent degree of APC activation as that produced by T_H cells, either through direct infection or provoking inflammatory host responses^{13,14}. Although this model provides a possible explanation for the conditional nature of T-cell help for CTL responses, it does not address which aspects of CD8⁺ development (for example, primary expansion, functional differentiation, or memory) are regulated by CD4⁺ T cells. We have now investigated this question for both T_H-dependent and T_H-independent CTL responses.

Immunization of C57BL/6 mice with syngeneic human adenovirus type 5 E1-transformed *Tap*^{−/−} mouse embryo cells (*Tap*^{−/−} 5E1 MECs) induces a prototypical T_H-dependent, CD8⁺ T-cell response through cross-priming¹⁵. Depletion of CD4⁺ cells before immunization blocks the development of E1B(192–200)-peptide-specific CTLs after secondary *in vitro* re-stimulation 14 days later³. To investigate whether the absence of T_H cells abrogates primary expansion, the frequency of interferon-γ (IFN-γ)⁺ effector CD8⁺ T cells was monitored directly *ex vivo* at weekly intervals after immunization of intact (wild type) or CD4-depleted mice. The

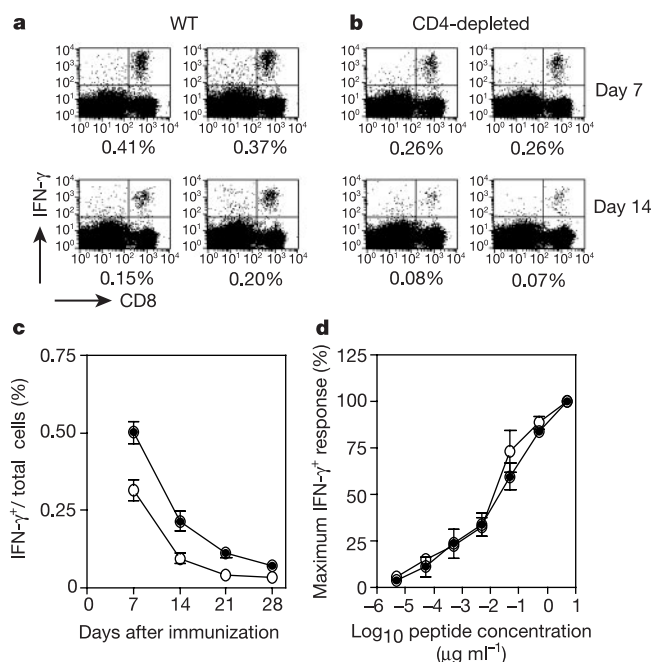


Figure 1 Functional expansion of T-helper (T_H) cell-dependent CD8⁺ T cells cross-primed in the absence of CD4⁺ T-cell help. **a, b**, The frequency of interferon-γ (IFN-γ)⁺ E1B(192–200)-specific CD8⁺ T cells detected in the spleens of *Tap*^{−/−} Ad5E1-immunized normal (WT) (**a**) or CD4-depleted (**b**) mice on stimulation with E1B(192–200) peptide. (Control peptide-stimulated responses were <0.01% of total splenocytes.) The value under each panel represents the percentage of total cells that are positive for IFN-γ. **c**, Percentage of IFN-γ⁺ E1B(192–200)-specific CD8⁺ cells in the spleen of normal (filled circles) and CD4-depleted (open circles) mice at various time points after immunization. **d**, Comparable functional avidity of E1B(192–200)-specific IFN-γ⁺ CD8⁺ cells of normal (filled circles) and CD4-depleted (open circles) mice after stimulation with titrated concentrations of E1B(192–200) peptide. Results are displayed as mean ± s.e.m. (*n* = 6) and are representative of three experiments.

absence of $CD4^+$ cells reduces the number of effector $CD8^+$ T cells detected at the day 7 peak of the response by approximately half compared with wild-type mice, and results in a lower per-cell level of $IFN-\gamma$ production (Fig. 1a, b; see also Supplementary Fig. S1). Despite the difference in peak burst sizes, the rate of subsequent contraction in the two antigen-specific effector populations did not differ significantly over the next 3 weeks (Fig. 1c). Similar data were obtained for E1B-specific $CD8^+$ T cells primed in $I-A\beta^{-/-}$ mice, which lack T_H cells because their APCs do not express major histocompatibility complex (MHC) class II molecules (data not shown)¹⁶. The reduction in peak burst size is not due to differences in the antigenic responsiveness of the CTLs primed in the presence versus the absence of T_H cells, as the $IFN-\gamma^+$ primary effectors in both populations displayed similar functional avidity for their peptide ligand (Fig. 1d; see also Supplementary Fig. S2). Taken together, these data show that ' T_H -dependent' $CD8^+$ T cells can be functionally primed in the absence of T_H cells to generate $IFN-\gamma$ -producing effectors.

The *in vivo* cytotoxic capacity of $CD8^+$ T cells primed in the absence of T_H cells was assessed within intact mice. This was done by monitoring the specific eradication of an adoptively transferred target population of E1B(192–200)-pulsed splenocytes that had been differentially labelled with carboxy fluorescein succinimidyl ester (CFSE^{high}) so as to be distinguishable from a co-transferred reference population (CFSE^{low}) pulsed with a control peptide¹⁷. Figure 2a shows that strong and specific cytotoxic effectors are detectable *in vivo* 7 and 14 days after priming in the absence of T_H

cells. These primary CTLs appear to have a transient capacity to generate secondary effectors after re-stimulation at day 7, but not at day 14 (Fig. 2b). In contrast, CTLs primed in the presence of T_H cells are detectable in secondary cultures at days 7 and 14 (Fig. 2b), as well as for several months after priming in wild-type mice (not shown). These results indicate that $CD4^+$ cells are not required for the development of T_H -dependent $CD8^+$ T cells into primary CTLs *in vivo*. The capacity of these CTLs to develop into secondary effectors detectable *in vitro*, however, seems to depend on the length of the interval between priming and re-stimulation, being present at day 7 but absent at day 14 after immunization.

The discrepancy between the presence of primary compared with secondary cytotoxic effectors at days 7 and 14 after immunization

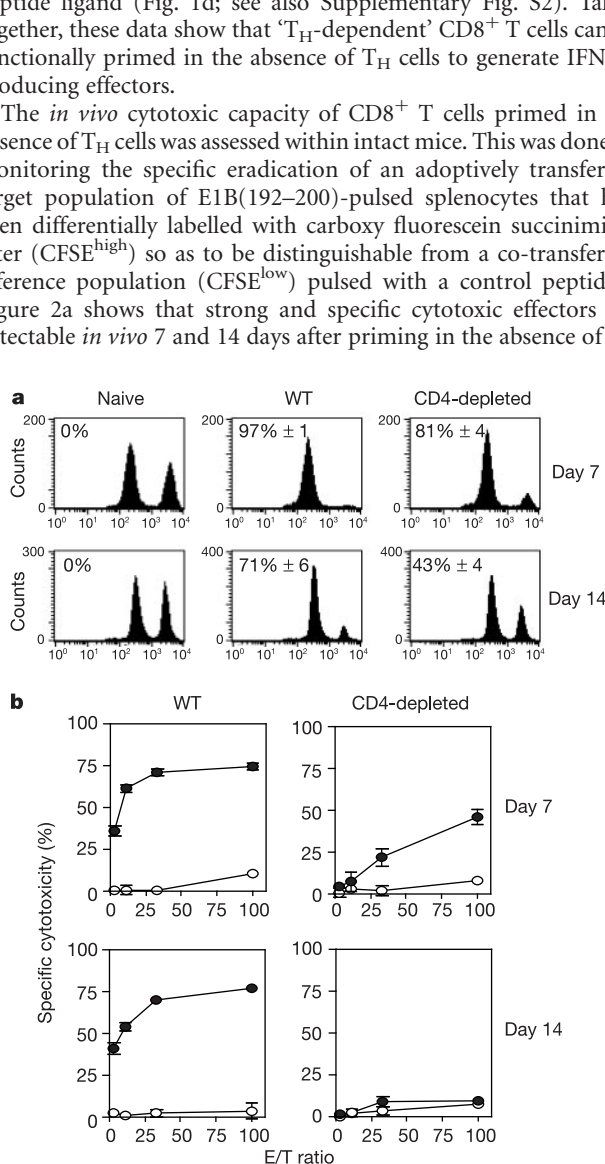


Figure 2 Primary and secondary cytotoxic responses of T_H -dependent cytotoxic T lymphocytes (CTLs). **a**, *In vivo* killing of E1B(192–200)-pulsed target cells labelled with carboxy fluorescein succinimidyl ester (CFSE^{high}) in normal or $CD4$ -depleted mice at various times after immunization with Tap^{-/-} 5E1 mouse embryo cells (MECs). The values in the left corner of each panel represent the percentage of specific killing compared with control-pulsed, CFSE^{low}-labelled cells (mean $\pm$ s.e.m.; $n = 6$). **b**, *In vitro* cytolytic activity of re-stimulated splenocytes from normal or $CD4$ -depleted mice 7 and 14 days after immunization. Various effector/target (E/T) ratios were tested for killing of syngeneic EL-4 cells pulsed with E1B(192–200) peptide (filled circles) or OVA(257–264) peptide (open circles). Values represent mean $\pm$ s.e.m. ($n = 6$).

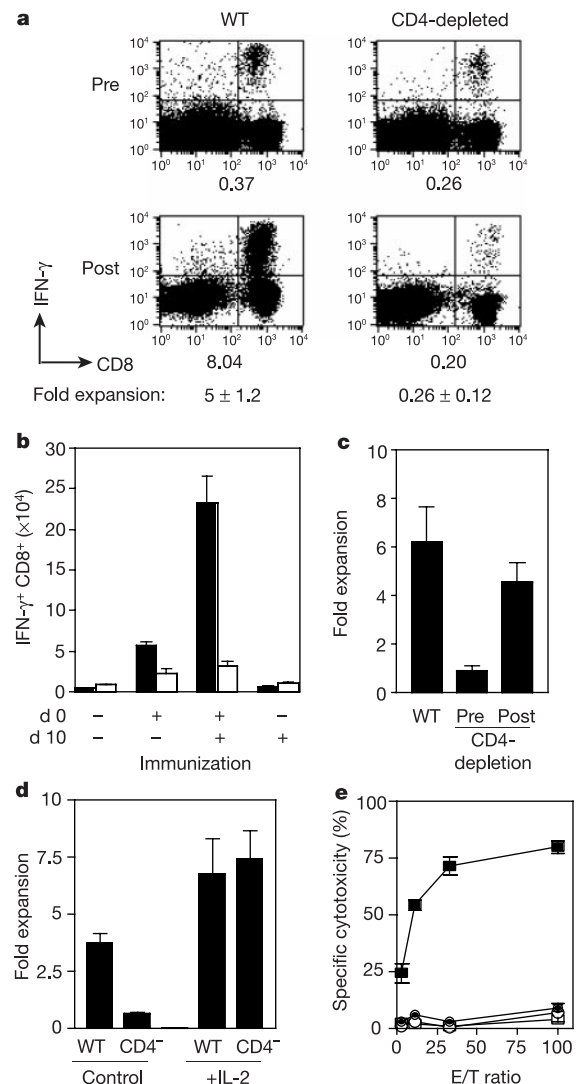


Figure 3 Defective secondary expansion of CTLs cross-primed in the absence of T_H cells. **a**, Per cent of $IFN-\gamma^+$ E1B(192–200)-specific CTLs among total splenocytes before and after *in vitro* re-stimulation. The value under each panel represents the per cent of total $CD8^+$ cells that are $IFN-\gamma^+$. Fold expansion is calculated from the increase in the absolute number of specific CTLs. **b**, Primary and secondary *in vivo* responses of E1B(192–200)-specific $IFN-\gamma^+$ CTLs detected directly in spleens of normal (filled bar) and $CD4$ -depleted (open bar) mice 4 days after secondary (d 10) immunization. **c**, Fold expansion after re-stimulation at day 7 of purified $CD8^+$ T cells from mice depleted of $CD4^+$ T cells before (pre) or after (post) immunization. **d**, Rescue of secondary expansion after re-stimulation by addition of interleukin-2 (IL-2). **e**, Cytotoxic activity of splenocytes from $CD4$ -depleted Tap^{-/-} 5E1 MEC-immunized mice after day 7 of re-stimulation. Squares, IL-2 added to re-stimulation culture; circles, no IL-2 added; filled symbols, EL-4 target pulsed with E1B(192–200) peptide; open symbols, EL-4 pulsed with control peptide.

prompted us to monitor CTL expansion after re-stimulation. This was accomplished by direct *ex vivo* enumeration of the antigen-specific IFN- γ ⁺ CTLs before *in vitro* re-stimulation and then again six days later. Figure 3a shows that, although a population of E1B-specific effector CTLs is readily detectable at day 7 after immunization in wild-type or CD4-depleted mice, only CTLs primed in the presence of T_H cells are able to expand (by fivefold on average) on secondary encounter with antigen. In contrast, CTLs primed in CD4-depleted mice fail to expand and generally contract after re-stimulation. This same pattern of secondary expansion of 'helped' CTLs from wild-type mice and contraction of 'helpless' CTLs from CD4-depleted mice seen at day 7 was also observed at 14, 21 and 28 days after immunization (Table 1).

Next, we investigated whether the helpless CTLs similarly failed to undergo secondary expansion *in vivo* by monitoring the number of E1B-specific IFN- γ ⁺ effector CTLs in wild-type or CD4-depleted mice that had received either a single immunization 14 days earlier (d 0), or a second immunization 10 days afterwards (d 10). Although E1B-specific primary effectors are induced in both wild-type and CD4-depleted mice, only CTLs primed in the presence of T_H cells are able to expand *in vivo* on re-challenge (Fig. 3b). We next investigated whether the presence of T_H cells during *in vitro* re-stimulation was required for secondary expansion of helped CTLs. This was accomplished by depleting CD4⁺ cells either before or after immunization and monitoring secondary expansion of purified CD8⁺ CTLs following re-stimulation. Whereas depletion of CD4⁺ cells before priming yields primary effector CTLs that fail to undergo secondary expansion, depletion of CD4 cells three days after priming has no such effect (Fig. 3c). These results indicate that the capacity for secondary expansion is programmed and becomes cell-autonomous after priming in the presence of T_H cells. The helpless CTLs are not irretrievably consigned to defective secondary expansion and cytotoxicity, however, as providing exogenous interleukin-2 (IL-2) at re-stimulation restores both of these functions (Fig. 3d, e). Taken together, these data demonstrate that T_H cells are required for secondary expansion of CTLs *in vitro* and *in vivo*, and that CTLs are endowed with this capacity during priming.

Our combined results indicate that for T_H-dependent responses, CD4⁺ cells are required to confer secondary, rather than primary, CTL expansion. To investigate whether this is also true for prototypical T_H-independent CTL responses, we monitored the primary and secondary CD8⁺ responses induced by infection with lymphocytic choriomeningitis virus (LCMV). Wild-type, CD4-depleted, or I-A β ^{-/-} mice were infected with LCMV (Armstrong), and the resulting LCMV GP(33–41)-peptide-specific CTL response was quantified physically, using D^b/GP(33–41) tetramers, and functionally, using intracellular detection of IFN- γ by GP(33–41)-specific D^b- and K^b-restricted CTLs^{18,19}. Although the magnitude of the GP(33–41)-specific response and cytotoxic function between the three groups was equivalent when measured 7 days after infection (data not shown), the GP(33–41)-specific CTLs primed in mice lacking T_H cells were reduced in number and produced less IFN- γ than wild-type controls when measured at day 28 (Fig. 4a, b). Nonetheless, both the I-A β ^{-/-} and CD4-depleted mice contained

a large number of antigen-specific CTLs at day 28, of the order of 1–1.5% of total CD8⁺ cells. On re-stimulation by LCMV-infected or GP(33–41)-pulsed peritoneal I-A β ^{-/-} macrophages expressing high levels of CD80 and CD86 (data not shown), only CTLs primed in the presence of T_H cells are able to undergo secondary expansion (Fig. 4c, d; see also Supplementary Fig. S3). These results demonstrate that, as with T_H-dependent responses, T_H-independent LCMV-specific CTLs primed in the absence of CD4⁺ cells also fail to expand on re-encounter with antigen.

Our results provide new insight into the nature of T-cell help by demonstrating that T_H cells are absolutely required for secondary, but not primary, CTL expansion. The progressive differentiation of lymphocytes involving initial CD4-independent and CD4-dependent events is reminiscent of the function performed by T_H cells in endowing B cells with the capacity for isotype switching²⁰. When considered in terms of how T_H-dependent CTLs have traditionally been defined (that is, as lytic effectors after *in vitro* re-stimulation) our findings can explain a number of previous observations, including why the requirement for T_H cells has traditionally been seen as 'conditional'. When primed in the absence of T_H cells, antigens that induce a modest CD8 response will fail to generate cytotoxic effectors detected through standard *in vitro* cytotoxicity assays, which rely on secondary expansion during re-stimulation. Antigens able to induce a large clonal expansion, in contrast, will yield CTL responses that are detectable directly *ex vivo* as well as in secondary cultures. Exceptions to this, as our study reveals, occur when a lymphocyte population containing a sufficiently large number of primary effectors are placed into *in vitro* culture. These can survive during re-stimulation and, despite failing to undergo secondary expansion, can nonetheless exhibit activity in subsequent *in vitro* cytotoxicity assays (Fig. 2a, b). This finding offers an explanation for the apparent dependence of transgenic CTLs on T-cell help on their precursor frequency in adoptive transfer experiments²¹. Conversely, the modest (average fivefold) expansion of re-stimulated CTLs and the relatively high sensitivity threshold of *in vitro* lysis assays may explain why CD8⁺ T cells primed in the presence of T_H cells may nonetheless fail to be detected months after primary antigenic challenge¹². The fact that added IL-2 was able to mask the secondary expansion defect in primary effectors suggests that re-stimulation protocols that feature addition of this

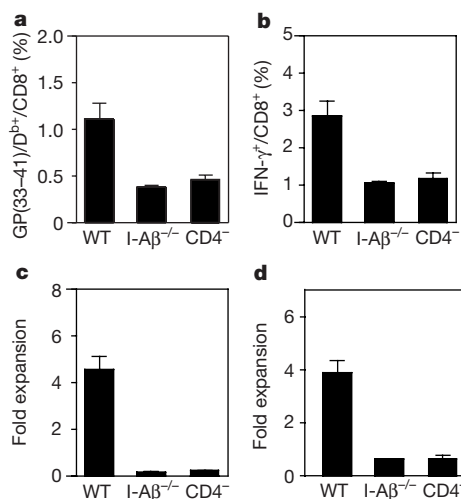


Figure 4 Defective secondary expansion of T_H-independent CD8⁺ T cells primed in the absence of CD4⁺ T-cell help. **a, b**, The frequency of GP(33–41)-specific T cells detected in total CD8⁺ splenocytes of mice 28 days after lymphocytic choriomeningitis virus (LCMV) infection as determined by GP(33–41)/D^b tetramers (**a**) and IFN- γ production (**b**) after stimulation with GP(33–41) peptide. **c, d**, The fold expansion of GP(33–41)-specific CD8⁺ T cells on secondary *in vitro* stimulation with I-A β ^{-/-} (MHC class II^{-/-}) macrophages that were infected with LCMV (**c**) or pulsed with GP(33–41) peptide (**d**).

Table 1 Secondary CTL expansion

Day	Fold expansion	
	WT mice	CD4-depleted mice
7	5.1 ± 0.9	0.3 ± 0.3
14	4.7 ± 0.8	0.3 ± 0.3
21	4.0 ± 1.0	0.1 ± 0.1
28	5.0 ± 1.9	0.1 ± 0.1

Interferon- γ -positive (IFN- γ ⁺), E1B(192–200)-specific cytotoxic T lymphocytes (CTLs) were monitored before and after *in vitro* re-stimulation of splenocytes from wild-type (WT) and CD4-depleted mice at various times after immunization with Tap^{-/-} 5E1 mouse embryo cells (MECs), as described in the Methods. The fold expansion is calculated from the increase in the absolute number of specific CTLs and represents the average of $n = 3$ mice. Similar results were observed in three separate experiments.

cytokine, including limiting dilution assays, may greatly overestimate the frequency of 'memory' CTLs primed under various conditions²². Furthermore, our results also suggest that the attrition of CTLs primed in the absence of T_H cells, as observed for LCMV-specific CTLs in Fig. 4a, b, may result from contraction after encounter with persistent antigen rather than an intrinsic lack of CTL survival²³.

The ability of peripheral T cells to expand on secondary encounter with antigen is a crucial element of immune memory and protective immunity^{5,6,24}. Our results indicate that helped CTLs are endowed with this capacity at a discreet step during their primary activation, and can express this memory phenotype as primary effectors as well as during and after their subsequent contraction phase. Our findings also suggest that a re-examination of the terms 'T_H-dependent' and 'T_H-independent' as applied to memory CTLs is needed, as many of the common methods used to monitor their physical or functional presence in short-term assays may instead detect primary effectors incapable of further clonal expansion^{22,25}. This may be especially important in clinical settings where accurate monitoring of 'memory phenotype' CTLs is desired, such as in therapeutic vaccination strategies. Finally, our data indicate that the instructional programme guiding CTL development is not invariant and can be modified to include or exclude specific functional capacities. Using T-cell help as a starting point, future efforts will be directed towards understanding how the signals that are transmitted to CD8 T cells are integrated into the particular differentiation programme executed. Such information may increase our ability to manipulate effectively CTL responses in infectious disease, cancer and autoimmunity. □

Methods

Mice and cell lines

C57BL/6J, B6.SJL-Ptpr^a (B6/SJL) and B6.SJL/I-A^b (all H-2^b) mice were purchased from The Jackson Laboratory. Mice were maintained by in-house breeding at the La Jolla Institute for Allergy and Immunology, and were maintained under specific pathogen-free conditions in accordance with guidelines by the Association for assessment and Accreditation of Laboratory Animal Care International.

Tap^{-/-} mouse embryo cell lines (MECs) expressing human adenovirus type 5 early region 1 (Ad5E1) were produced by transfection of both C57BL/6 Tap^{+/+} and Tap^{-/-} MEC lines and have been described previously¹⁵. The EL-4 thymoma and MC57 fibroblasts were purchased from the American Type Culture Collection. MECs were cultured in DMEM medium, and EL-4 and MC57 cells were cultured in IMDM medium. All media were supplemented with 10% fetal calf serum, 50 µM 2-mercaptoethanol, 2 mM L-glutamine, 20 U ml⁻¹ penicillin and 20 µg ml⁻¹ streptomycin.

Immunizations and antibody treatment

Mice were immunized subcutaneously in the right flank with 1 × 10⁷ irradiated (3,000 R) Tap^{-/-} 5E1 MECs, or were inoculated intraperitoneally with 1 × 10⁵ plaque-forming units of LCMV (Armstrong strain). Depletion of CD4⁺ cells *in vivo* was performed by intraperitoneal administration of 150 µg GK1.5 antibody on the first three days before immunization and every third day thereafter²⁶. In parallel experiments, GK1.5 was administered 3 days after immunization for 3 consecutive days.

Isolation of CD8⁺ T cells

CD8⁺ T cells were purified from the spleens and lymph nodes of previously immunized mice by antibody-directed complement lysis using antibodies to CD4 (GK1.5), class II MHC (M5/114, Y17 and CA-4.A12), B220 (RA3.6.B2), CD11b (M1/70), dendritic cells (33D1), natural killer cells (PK136) and heat-stable antigen (JIID), as described previously²⁷. The resulting cells were greater than 95% pure CD8⁺ cells, and contained less than 0.1% CD4⁺ T cells, as demonstrated by fluorescence-activated cell sorting (FACS) analysis.

In vitro re-stimulations

For E1B-specific responses, splenocytes (or purified CD8⁺ T cells) from Tap^{-/-} 5E1 MEC-immunized mice were collected at different time points after immunization, and stimulated for 6 days *in vitro* with irradiated syngeneic Tap^{+/+} Ad5E1 MECs (10:1 ratio). In parallel cultures, 10 ng ml⁻¹ IL-2 (Cetus) was added. For LCMV-specific responses, splenocytes derived from LCMV-infected mice were stimulated for 7 days *in vitro* at a 10:1 ratio with irradiated (3,000 R) syngeneic I-A^b thiolglycollate-induced macrophages infected with LCMV or pulsed with GP(33–41) peptide (KAVYNFATC; 0.5 µg ml⁻¹). At the end of all re-stimulation cultures, viable cells were collected by Ficoll gradient (Lympholyte-M; Cedarlane Laboratories).

Enumeration of specific CD8⁺ T cells

Directly *ex vivo*, after CD8 purification or after *in vitro* re-stimulation, cells were incubated for 5 h with either the E1B(192–200) peptide (VNIRNCCYI; 0.5 µg ml⁻¹), GP(33–40) (KAVYNFATC; 0.2 µg ml⁻¹) or OVA(257–264) (SIINFEKL; 0.2–0.5 µg ml⁻¹) in the

presence of brefeldin A. Surface staining for CD8 and intracellular cytokine staining for IFN-γ and tumour necrosis factor-α were performed using a Cytofix/Cytoperm Kit (Pharmingen) according to the manufacturer's directions. D^b/GP(33–41) tetramer staining was performed as described previously¹⁹. The fold expansion of specific CD8⁺ cells was calculated by dividing the absolute number of IFN-γ⁺ CD8⁺ cells after *in vitro* culture by the absolute number of IFN-γ⁺ CD8⁺ cells at the start of the culture.

Cytotoxicity assays

In vivo cytolytic activity was determined using B6.SJL spleen cells differentially labelled with the fluorescent dye CFSE¹⁷. The cells labelled with CFSE^{high} were used as targets and pulsed with E1B(192–200) peptide (0.5 µg ml⁻¹; 90 min at 37 °C, 5% CO₂), whereas the cells labelled with CFSE^{low} were pulsed with OVA(257–264) (0.5 µg ml⁻¹) to serve as the internal control. Peptide-pulsed target cells were extensively washed to remove free peptide and then co-injected intravenously in a 1:1 ratio to previously immunized B6 mice. Sixteen hours later, spleens were removed and the ratio of CFSE^{low}/CFSE^{high} cells was determined by flow cytometry. The values in the upper left corner of each panel represents the per cent loss of E1B(192–200)-pulsed (CFSE^{high}) target cells.

In vitro cytolytic activity of re-stimulated splenocytes of Ad5E1 MEC-immunized mice was evaluated by a JAM test as previously described²⁷, using [³H]thymidine-labelled EL-4 cells loaded with E1B(192–200) peptide or OVA(257–264) peptide. Specific killing was calculated as follows: (spontaneous c.p.m. – experimental c.p.m.) × 100/spontaneous c.p.m., where c.p.m. is counts per minute. LCMV-specific cytolytic activity was assessed in a standard 5-h [⁵¹Cr] release assay on LCMV-infected and uninfected MC57 cells, and plotted as a percentage of the maximum lysis after subtracting the nonspecific lysis (always less than 10%).

Data analysis

Unless stated otherwise, data are expressed as mean ± standard error of the mean, and evaluated using an analysis of variance followed by a Dunnett test. A probability value of *P* < 0.05 was considered statistically significant.

Received 3 December 2002; accepted 23 January 2003; doi:10.1038/nature01441.

Published online 9 February 2003.

1. Buller, R. M., Holmes, K. L., Hugin, A., Frederickson, T. N. & Morse, H. C. III Induction of cytotoxic T-cell responses *in vivo* in the absence of CD4 helper cells. *Nature* **328**, 77–79 (1987).
2. Rahemtulla, A. *et al.* Normal development and function of CD8⁺ cells but markedly decreased helper cell activity in mice lacking CD4. *Nature* **353**, 180–184 (1991).
3. Schoenberger, S. P., Toes, R. E., van der Voort, E. L., Offringa, R. & Melief, C. J. T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions. *Nature* **393**, 480–483 (1998).
4. Bennett, S. R., Carbone, F. R., Karamalis, F., Miller, J. F. & Heath, W. R. Induction of a CD8 + cytotoxic T lymphocyte response by cross-priming requires cognate CD4 + T cell help. *J. Exp. Med.* **186**, 65–70 (1997).
5. Sprent, J. & Surh, C. D. T cell memory. *Annu. Rev. Immunol.* **20**, 551–579 (2002).
6. Ahmed, R. & Gray, D. Immunological memory and protective immunity: understanding their relation. *Science* **272**, 54–60 (1996).
7. van Stipdonk, M. J., Lemmens, E. E. & Schoenberger, S. P. Naive CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nature Immunol.* **2**, 423–429 (2001).
8. Kaech, S. M. & Ahmed, R. Memory CD8 + T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. *Nature Immunol.* **2**, 415–422 (2001).
9. Badovinac, V. P., Porter, B. B. & Harty, J. T. Programmed contraction of CD8(+) T cells after infection. *Nature Immunol.* **3**, 619–626 (2002).
10. Keene, J. & Forman, J. Helper activity is required for the *in vivo* generation of cytotoxic T lymphocytes. *J. Exp. Med.* **155**, 768–782 (1982).
11. Bennett, S. R. *et al.* Help for cytotoxic-T-cell responses is mediated by CD40 signalling. *Nature* **393**, 478–480 (1998).
12. Ridge, J. P., Di Rosa, F. & Matzinger, P. A conditioned dendritic cell can be a temporal bridge between a CD4⁺ T-helper and a T-killer cell. *Nature* **393**, 474–478 (1998).
13. Matzinger, P. Tolerance, danger, and the extended family. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
14. Lanzavecchia, A. Immunology. Licence to kill. *Nature* **393**, 413–414 (1998).
15. Schoenberger, S. P. *et al.* Cross-priming of CTL responses *in vivo* does not require antigenic peptides in the endoplasmic reticulum of immunizing cells. *J. Immunol.* **161**, 3808–3812 (1998).
16. Grusby, M. J., Johnson, R. S., Papaioannou, V. E. & Glimcher, L. H. Depletion of CD4 + T cells in major histocompatibility complex class II-deficient mice. *Science* **253**, 1417–1420 (1991).
17. Oehen, S. & Brduscha-Riem, K. Differentiation of naive CTL to effector and memory CTL: correlation of effector function with phenotype and cell division. *J. Immunol.* **161**, 5338–5346 (1998).
18. Hudrisier, D., Oldstone, M. B. & Gairin, J. E. The signal sequence of lymphocytic choriomeningitis virus contains an immunodominant cytotoxic T cell epitope that is restricted by both H-2D(b) and H-2K(b) molecules. *Virology* **234**, 62–73 (1997).
19. Busch, D. H., Pilip, I. M., Vijh, S. & Pamer, E. G. Coordinate regulation of complex T cell populations responding to bacterial infection. *Immunity* **8**, 353–362 (1998).
20. Parker, D. C. T cell-dependent B cell activation. *Annu. Rev. Immunol.* **11**, 331–360 (1993).
21. Mintern, J. D., Davey, G. M., Belz, G. T., Carbone, F. R. & Heath, W. R. Cutting edge: precursor frequency affects the helper dependence of cytotoxic T cells. *J. Immunol.* **168**, 977–980 (2002).
22. Di Rosa, F. & Matzinger, P. Long-lasting CD8 T cell memory in the absence of CD4 T cells or B cells. *J. Exp. Med.* **183**, 2153–2163 (1996).
23. Klennerman, P., Hengartner, H. & Zinkernagel, R. M. A non-retroviral RNA virus persists in DNA form. *Nature* **390**, 298–301 (1997).
24. Kaech, S. M., Wherry, E. J. & Ahmed, R. Effector and memory T-cell differentiation: implications for vaccine development. *Nature Rev. Immunol.* **2**, 251–262 (2002).
25. Suresh, M. *et al.* Role of CD28-B7 interactions in generation and maintenance of CD8 T cell memory. *J. Immunol.* **167**, 5565–5573 (2001).
26. Dialynas, D. P. *et al.* Characterization of the murine T cell surface molecule, designated L3T4, identified by monoclonal antibody GK1.5: similarity of L3T4 to the human Leu-3/T4 molecule. *J. Immunol.* **131**, 2445–2451 (1983).

27. Matzinger, P. The JAM test. A simple assay for DNA fragmentation and cell death. *J. Immunol. Methods* **145**, 185–192 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by grants from the National Institutes of Health, the American Cancer Society, and CaP CURE to S.P.S. The authors thank G. Palioungas for technical assistance and the staff of the LIAI animal facility for animal care.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.P.S. (e-mail: sps@liai.org).

STAT3 signalling is required for leptin regulation of energy balance but not reproduction

Sarah H. Bates*, Walter H. Stearns*, Trevor A. Dundon*, Markus Schubert*, Annette W. K. Tso*, Yongping Wang†, Alexander S. Banks*, Hugh J. Lavery*, Asma K. Haq*, Eleftheria Maratos-Flier*, Benjamin G. Neel†, Michael W. Schwartz‡ & Martin G. Myers Jr*

* Research Division, Joslin Diabetes Center, and † Cancer Biology Program, Beth Israel Deaconess Medical Center and Harvard Medical School, 1 Joslin Place, Boston, Massachusetts 02215, USA

‡ Harborview Medical Center, University of Washington, Seattle, Washington 98122, USA

Secretion of leptin from adipocytes communicates body energy status to the brain by activating the leptin receptor long form (LRb). LRb regulates energy homeostasis and neuroendocrine function; the absence of LRb in *db/db* mice results in obesity, impaired growth, infertility and diabetes^{1–4}. Tyr 1138 of LRb mediates activation of the transcription factor STAT3 during leptin action^{5–8}. To investigate the contribution of STAT3 signalling to leptin action *in vivo*, we replaced the gene encoding the leptin receptor (*lepr*) in mice with an allele coding for a replacement of Tyr 1138 in LRb with a serine residue (*lepr*^{S1138}) that specifically disrupts the LRb–STAT3 signal. Here we show that, like *db/db* mice, *lepr*^{S1138} homozygotes (*s/s*) are hyperphagic and obese. However, whereas *db/db* mice are infertile, short and diabetic, *s/s* mice are fertile, long and less hyperglycaemic. Furthermore, hypothalamic expression of neuropeptide Y (NPY) is elevated in *db/db* mice but not *s/s* mice, whereas the hypothalamic melanocortin system is suppressed in both *db/db* and *s/s* mice. LRb–STAT3 signalling thus mediates the effects of leptin on melanocortin production and body energy homeostasis, whereas distinct LRb signals regulate NPY and the control of fertility, growth and glucose homeostasis.

To test the requirement for LRb–STAT3 signalling in the physiological functions of leptin, we generated the *lepr*^{S1138} mutant by homologous gene targeting (Fig. 1a). LRb^{S1138}, the protein encoded by *lepr*^{S1138}, fails to activate STAT3 although it is expressed normally (Fig. 1b, c). We used these targeted 129Sv embryonic stem cell clones to produce chimaeric animals, which were crossed with C57Bl/6 mice to generate heterozygous ‘knock-in’ mice expressing LRb^{S1138} in a manner identical to wild-type LRb. Interbreeding of heterozygous animals yielded wild-type (+/+), heterozygous (*s/+*) and homozygous (*s/s*) mice for study.

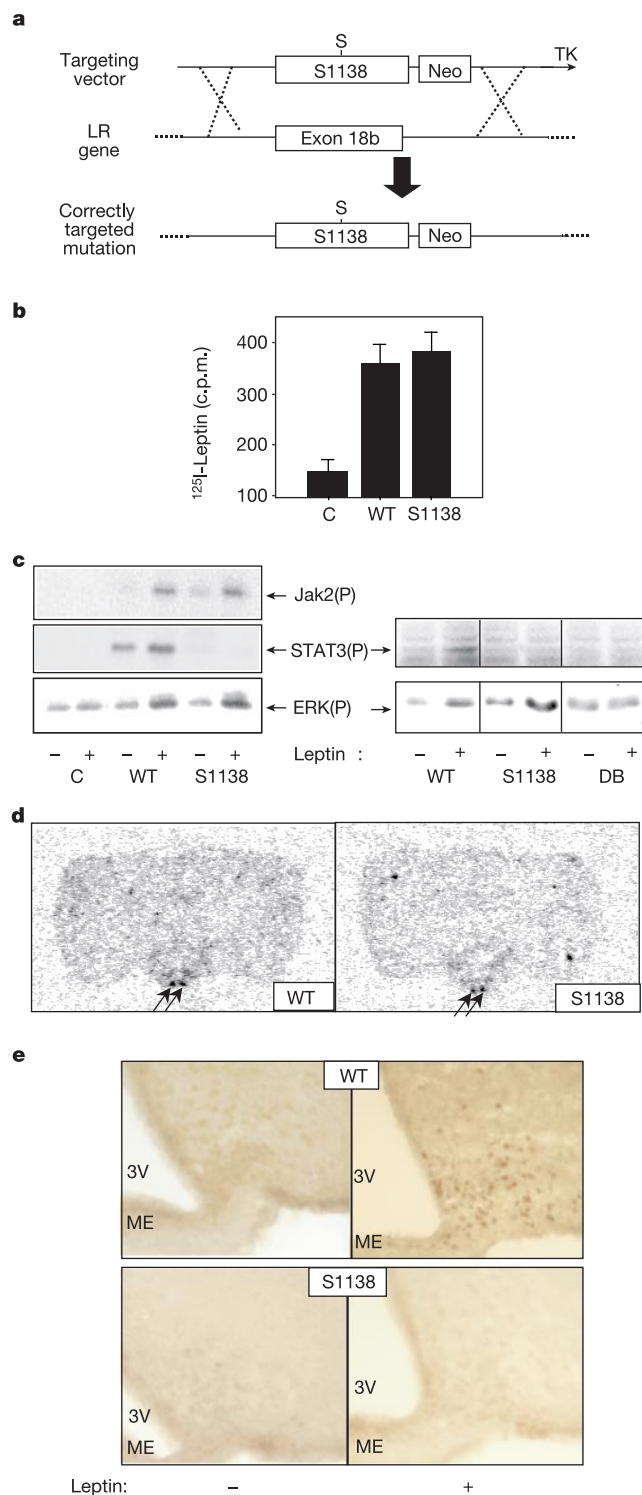


Figure 1 Generation of mice expressing LRb^{S1138}. **a**, Diagram of gene-targeting strategy to replace the LRb-specific exon 18b with the mutant exon S1138 (S; Tyr 1138 → Ser). **b**, Cell-surface ¹²⁵I-leptin binding by LRb^{S1138} in COS-7 cells (C), expressing LRb (WT) or LRb^{S1138} (S1138). **c**, Signalling by LRb^{S1138} in HEK-293 cells (left panels) and hypothalamus (right panels). Shown are STAT3, Jak2 and ERK phosphorylation after treatment with leptin (+) or vehicle (-). **d**, Distribution of LRb and LRb^{S1138} expression by *in situ* hybridization with LRb-specific probes (see Supplementary Information). LRb probe hybridization is indicated (arrows). **e**, Leptin-stimulated nuclear translocation of STAT3 in the hypothalamus. Hypothalamic sections from wild-type or S1138 mice treated with saline (-) or leptin (+) were immunostained for STAT3 (see Supplementary Information). Nuclear STAT3 is visible as distinct darkly staining nuclei. Third ventricle (3V) and median eminence (ME) are indicated. Mice: +/+ (wild type; WT), *s/s* (S1138), *db/db* (DB).

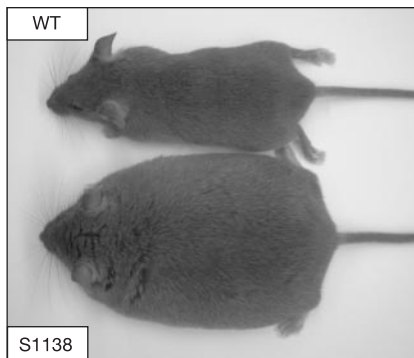


Figure 2 Obese phenotype of *s/s* mice. Shown is a photograph of littermate (C57;129) wild-type (WT) and S1138 mice at 10 weeks of age.

To confirm that the replacement of *lepr* by *lepr*^{S1138} yields LRB^{S1138} expression comparable to wild-type LRB, we documented similar levels of LRB and LRB^{S1138} messenger RNA expression by semiquantitative polymerase chain reaction with reverse transcription (RT-PCR) of hypothalamic mRNA using exon-spanning primers (data not shown). *In situ* hybridization histochemistry showed a normal distribution of mLRb mRNA in the mediobasal hypothalamus of *+/+* and *s/s* mice (Fig. 1d) (see Supplementary Information). Furthermore, as we have previously shown for *ob/ob* mice⁹, leptin stimulated the activation of extracellular signal-regulated kinase (ERK) in the hypothalamus of *+/+* and *s/s* animals but not in LRB-deficient *db/db* mice (Fig. 1c). In contrast, leptin failed to induce STAT3 tyrosine phosphorylation and nuclear translocation in *s/s* animals (Fig. 1c, e), whereas these effects were clearly detected in *+/+* mice. Thus, although LRB^{S1138} can activate non-STAT3 neuronal signals *in vivo*, it is incapable of STAT3 activation.

Obesity developed at an early age in *s/s* mice as in *db/db* mice (Fig. 2); these animals ultimately attained weights twofold to threefold those of wild-type mice. To facilitate comparisons between *s/s* mice (lacking only the LRB → STAT3 signal) and *db/db* mice (which lack LRB and thus all LRB signals), we used *s/s* and *db/db* animals on the C57BL/6 congenic background for our subsequent phenotypic analysis (Table 1). This comparison demonstrated that, although severely obese, *s/s* males were about 10% lighter than age-matched *db/db* controls (Table 1). This disparity was more pronounced in *s/s* females, which were 20% lighter than *db/db* females. The increased body weights of *s/s* animals, as for *db/db* animals, reflected increased whole-body triacylglycerol levels (data not shown). As in *db/db* mice, obesity in *s/s* mice was characterized by hyperphagia despite markedly elevated serum leptin levels, indicating severe leptin resistance (Table 1). Because *s/s* mice paired to the intake of wild-type mice gained more weight and adipose mass than their wild-type controls (data not shown), reduced energy expenditure also seems to be a feature of the *s/s* phenotype, as for *db/db* mice¹⁰. The LRB-STAT3 signal is therefore central to the regulation of food intake and energy expenditure by leptin. Furthermore, the amplitude of the LRB-STAT3 signal seems critical

to this process, as *s/+*, like *db/+*, heterozygotes weigh slightly more than wild-type mice.

Commensurate with their level of obesity, *s/s* and *db/db* mice were similarly hyperinsulinaemic (Table 1); however, lower blood glucose levels were observed in *s/s* mice than in *db/db* mice, indicating improved glycaemic control in *s/s* mice. Taken together, these data indicate that although LRB-STAT3 signals contribute to leptin action on both energy homeostasis and glucose homeostasis, LRB-STAT3 signalling is more important for the former, whereas STAT3-independent signals are more critical to the latter.

Starvation activates energy-sparing neuroendocrine responses that blunt the hypothalamic regulation of growth and gonadal function. Reduced leptin signalling mediates an important component of this starvation response, and an absence of leptin action results in infertility (especially among females, in which the energetic demands of reproduction are very high) and decreased linear growth in mice, rats and humans^{11–15}. The phenotype of *s/s* mice suggests a major role for STAT3-independent leptin signals in these neuroendocrine responses (Table 1). An analysis of linear growth by measurement of snout-anus length demonstrated that, whereas *db/db* mice were about 5% shorter than *+/+* controls, *s/s* mice were about 5% longer than controls (Table 1). Thus, the growth axis seems to be activated by the selective loss of the LRB-STAT3 signal in *s/s* mice, whereas the abrogation of all leptin signals in *db/db* mice inhibits normal growth.

Breeding studies with wild-type mates demonstrated that, whereas *db/db* females failed to reproduce, about 40% of female *s/s* mice were fertile (Table 1). Because reproduction is determined not only by gonadal function but also by behavioural and other complex inputs, we measured gonad function separately from behaviour by examining vaginal oestrogenization (Table 1). Female *db/db* mice never exhibited signs of vaginal oestrus, whereas *s/s* females displayed normal vaginal oestrogenization with only a slight delay in the onset of oestrous cycling compared with wild-type mice (median onset at 42 versus 36 days of age), and all *s/s* females demonstrated oestrous cycling. Moreover, whereas *db/db* females had atrophic reproductive organs, the organs of *s/s* females seemed grossly normal and all *s/s* ovaries revealed histological evidence of ovulation with normal numbers of corpora lutea (data not shown). Thus, in contrast to the absent ovulatory function of *db/db* females, gonadal function in all *s/s* females is essentially normal. The hypothalamic control of reproduction by leptin is therefore predominantly regulated by signals independent of STAT3 signalling and, like the regulation of glucose homeostasis and linear growth, depends on signals that are unperturbed in LRB^{S1138}.

The offspring of fertile *s/s* females died within 24 h of birth without milk in their stomachs unless fostered to wild-type mothers, suggesting a defect in lactation that was independent of gonad function in *s/s* females. Others have observed that withdrawal of leptin therapy from pregnant *ob/ob* females that had been leptin-treated to enable reproduction results in a similar outcome, indicating a role for leptin in the hypothalamic regulation of lactation¹⁶; our data suggest a role for the LRB → STAT3 signal in the control of lactation¹⁷.

Table 1 Phenotypic data for mice carrying *lepr*^{S1138} or *lepr*^{db}

Genotype	<i>+/+</i>	<i>s/+</i>	<i>s/s</i>	<i>db/+</i>	<i>db/db</i>
Weight (male) (g)	23.4 ± 0.4	24.2 ± 0.5	41.9 ± 0.9**†††	26.2 ± 0.9*	47.2 ± 2.7**††
Weight (female) (g)	19.6 ± 0.2	20.9 ± 0.6*	37.6 ± 0.6**†††	22.0 ± 0.6**	46.0 ± 1.4**††
Leptin (ng ml ⁻¹)	3.4 ± 1.1	3.8 ± 1.1	56.8 ± 5.5**†††	3.5 ± 1.4	67.7 ± 3.5††
Feeding (g per animal)	3.6 ± 0.2	3.4 ± 0.3	7.9 ± 0.7**††	3.9 ± 0.3	9.4 ± 0.9**††
Insulin (ng ml ⁻¹)	1.4 ± 0.3	2.1 ± 0.8	28 ± 4**††	1.2 ± 0.3	20 ± 2**††
Glucose (mM)	8.5 ± 0.4	7.2 ± 0.4	14.0 ± 1.5**†††	8.1 ± 0.4	29.7 ± 1.3**††
Snout-anus length (mm)	90 ± 1	(ND)	95 ± 1**††	(ND)	84 ± 2**
Fertility (F)	9/10	(ND)	3/7	(ND)	0/10
Onset of first oestrous cycle (d)	35	(ND)	42	(ND)	>98 (never)

Student's unpaired *t*-test: * *P* < 0.05 versus *+/+*; ** *P* < 0.001 versus *+/+*; † *P* < 0.05 versus heterozygote; †† *P* < 0.001 versus heterozygote; ‡ *P* < 0.05 versus *db/db*; ‡‡ *P* < 0.001 versus *db/db*. Differences not statistically significant unless noted. ND, not determined.

The phenotype of *s/s* animals cannot be attributed to a gross disturbance of LRB mRNA expression or hypothalamic architecture, because LRB mRNA expression (Fig. 1d) is intact in *s/s* mice, and thionin-stained hypothalamic sections reveal no anatomical abnormalities (data not shown). To investigate hypothalamic mechanisms that contribute to differences in the phenotypes of *s/s* and *db/db* mice, we studied leptin-regulated hypothalamic neuropeptides. Two distinct populations of LRB-expressing neurons in the arcuate nucleus of the hypothalamus (a key brain area for transducing input from leptin into behavioural and neuroendocrine responses) mediate important elements of leptin action^{10,18}. Neurons that produce the melanocortin peptide α -melanocyte-stimulating hormone (α -MSH) (which is cleaved from the pro-opiomelanocortin (POMC) precursor), mediate signals by the activation of neuronal melanocortin-3 and -4 receptors (MC3R, MC4R)^{10,18,19}. Leptin-induced anorexia depends at least in part on the activation of these LRB/POMC neurons²⁰. An adjacent sub-

population of LRB-expressing neurons synthesize the appetite-enhancing (orexigenic) peptides NPY and agouti-related protein (AgRP, an antagonist of MC3R and MC4R), and are inhibited by leptin^{18,20–22}. In addition to its orexigenic effects, NPY can also inhibit growth and reproductive function. Such effects are observed both after continuous central NPY infusion in normal rodents and in models with genetic obesity due to defective leptin signalling, including *ob/ob* and *db/db* mice^{15,23}. Indeed, the *s/s* phenotype resembles that displayed by *ob/ob* mice rendered NPY-deficient by *Npy* gene knockout (*ob/ob,Npy*^{-/-}) mice. These mice have attenuated obesity, increased body length and partly restored reproductive function compared with fatter, infertile, leptin-deficient *ob/ob* mice¹⁵. Also like *ob/ob,Npy*^{-/-} mice, *s/s* mice display improved grooming compared with *ob/ob* and *db/db* mice (data not shown) and a more marked improvement in energy homeostasis in females than in males (Table 1)¹⁵.

We thus proposed that the LRB–STAT3 signal mediates melanocortin function by leptin but does not regulate hypothalamic NPY expression. Accordingly, mice with selective impairment of the LRB–STAT3 signal should have low hypothalamic POMC mRNA levels and should therefore manifest hyperphagia and obesity, but unlike *db/db* mice they should not have elevated NPY mRNA and should retain intact growth and reproduction function. We therefore measured hypothalamic POMC, AgRP and NPY mRNA in *s/s* and *db/db* mice by RT–PCR (Fig. 3a–c) (see Supplementary Information). In *s/s* mice, as in *db/db* mice, the expression of hypothalamic POMC mRNA was reduced and that of AgRP mRNA was increased compared with control mice, confirming that STAT3 is required for the regulation of both mRNA species by LRB. This combination of decreased POMC and increased AgRP gene expression suggests an impairment of hypothalamic melanocortin action in both *s/s* and *db/db* mice, a mechanism that probably has a key role in the obese phenotype of *s/s* and *db/db* mice. However, whereas hypothalamic NPY mRNA levels were increased in *db/db* mice, NPY mRNA expression in *s/s* mice was similar to that in control mice. The finding of less pronounced AgRP mRNA overexpression in *s/s* mice than in *db/db* mice indicates that leptin might regulate AgRP expression by both STAT3-independent and STAT3-dependent mechanisms. This discordance in the molecular mechanisms regulating hypothalamic levels of NPY and AgRP mRNA is intriguing, because the two peptides are co-expressed in the same cells and are controlled in parallel by leptin and nutritional status in normal mice²⁴. Taken together, these findings suggest that whereas LRB–STAT3 signalling is required for both the stimulation of POMC and the full repression of AgRP transcription and thus normal melanocortin signalling, leptin-induced inhibition of NPY gene expression is mediated by a distinct signal. Indeed, the increased linear growth and intact reproductive function displayed by *s/s* mice are reminiscent of other mouse models with impaired melanocortin signalling (*A^y*, *Mc4r*^{-/-}, *ob/ob,Npy*^{-/-}) that are characterized by obesity and the normalization or repression of NPY expression¹⁵.

It is possible that SH2-domain-containing proteins other than STAT3 interact with Tyr 1138 of LRB and that the phenotype of *s/s* mice reflects defective signalling by molecules other than STAT3. However, in the absence of any such candidate molecule we attribute the phenotype of *s/s* mice to a disruption of LRB–STAT3 signalling. Although it is formally possible that some alteration in LRB^{S1138} protein level or trafficking might contribute to the phenotype of *s/s* animals, the normal trafficking and cell-surface expression of LRB^{S1138} in cultured cells (Fig. 1 and data not shown) and the ability of LRB^{S1138} to mediate ERK activation in the hypothalamus (Fig. 1) indicate that such alterations are not responsible for the observed phenotypes. Hence, STAT3-independent signals triggered by LRB seem to be especially important in the control of NPY expression, glucose homeostasis, linear growth, and fertility. Candidate LRB–STAT3 independent signals include ERK and PtdIns-3-OH kinase (PI(3)K)^{19,21,25}. Indeed, the regulation of

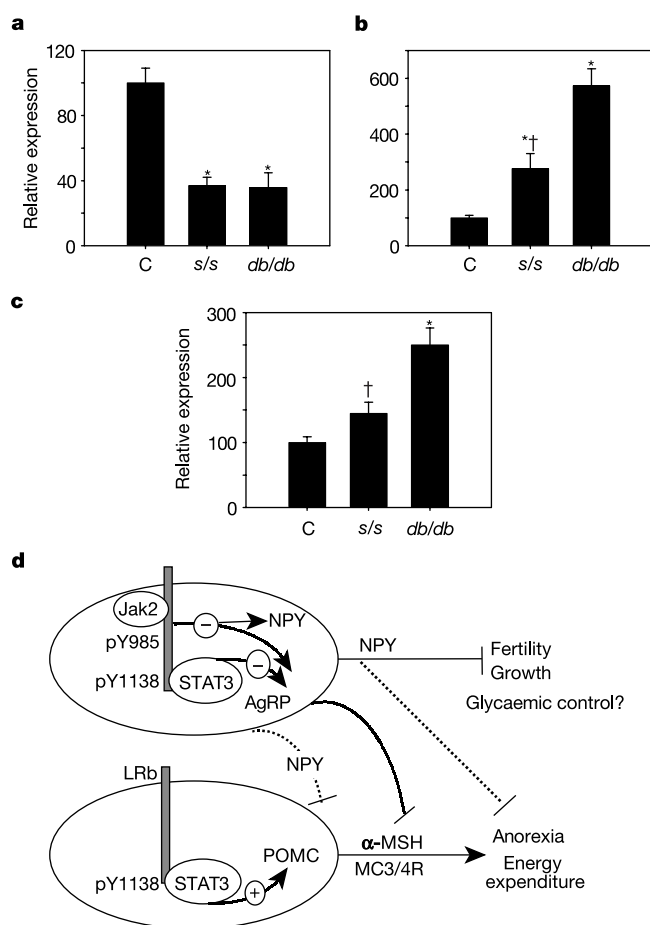


Figure 3 Hypothalamic neuropeptide mRNA levels and the control of physiology in *s/s* and *db/db* mice. **a–c**, Quantification of hypothalamic POMC (**a**), AgRP (**b**) and NPY (**c**) mRNA by automated fluorescent RT–PCR (see Supplementary Information) of RNA from 8-week-old C57Bl/6 control (C), *s/s* and *db/db* mice fed *ad libitum*. Results are normalized to a control value of 100%. Student's unpaired *t*-test: asterisk, *P* < 0.01 versus control; dagger, *P* < 0.01 versus *db/db*. Differences are not statistically significant unless noted. **d**, Model of LRB signals in the control of neuropeptide levels and physiology in arcuate nucleus. Leptin increases POMC production via STAT3, generating a positive anorectic signal by means of α -MSH and the melanocortin MC3 and MC4 receptors. In the LRB/NPY/AgRP neuron, leptin inhibits AgRP production in part by means of the LRB–STAT3 pathway, disinhibiting (potentiating) melanocortin signalling. Elevated NPY inhibits growth and reproductive axes and might influence glycaemic control. NPY also blocks a component of the anorectic function of leptin (dashed lines) directly and/or by cross-talk to the melanocortin system¹⁹. The inhibition of NPY by LRB proceeds independently of STAT3 signalling.

membrane potential in arcuate (presumably NPY) neurons by leptin is too rapid to be mediated by STAT3-driven transcription and requires PI(3)K activity^{19,21}.

Understanding the role of individual leptin signals in the control of physiology will enable the identification of sites of leptin resistance and potential therapeutic targets in obesity. Although the disruption of LRB-STAT3 signalling in *s/s* mice results in resistance to the effects of leptin in regulating energy balance, *s/s* mice display increased linear growth, intact reproductive function, and impaired glycaemic control compared with mice with globally disrupted leptin action. Similarly, the leptin resistance that accompanies diet-induced obesity is characterized by the preservation of growth and reproductive capabilities. Our findings therefore indicate that the pathogenesis of leptin resistance in common forms of obesity might involve attenuated LRB-STAT3 signalling. □

Methods

Animals

Interbreeding of *s/+* animals yielded *+/+*, *s/+* and *s/s* mice for study at the expected mendelian frequencies. C57Bl/6 *db/+* and *db/db* animals for breeding and experimentation were from Jackson Laboratories. Congenic C57Bl/6 *s/+* animals used to generate *s/s* mice were *n* = 6 generations backcrossed; genotyping a panel of 88 microsatellite markers revealed 99–100% homozygosity for C57Bl/6 alleles at these loci with the exception of a congenic segment surrounding *lepr* on chromosome 4 (98 to >99% homozygous for C57Bl/6 including the *lepr* segment). All mice were housed in the Joslin Diabetes Center's accredited mouse barrier facility with access to standard chow and water *ad libitum*, except those mice that were pair-fed to wild-type mice (given access to water *ad libitum* and fed the amount of chow eaten the previous day by a wild-type littermate) and those fed to wild-type weight as described below. Except for mice in breeding studies, which were housed in pairs, all mice were housed singly from the time of weaning at 21 days. All procedures were approved by the Joslin Diabetes Center Institutional Animal Care and Use Committee.

Phenotypic data

For measurement of the data shown in Table 1, male mice of the indicated genotype were weighed and killed between 10:00 and 12:00 during their eighth week of life. Blood was collected for glucose measurements with a glucometer and for serum determination of insulin and leptin (ELISA; Crystal Chem). Snout–anus length was measured with a micrometer on 8-week-old anaesthetized animals. Food consumption in 8-week-old male mice was determined by weighing before the dark cycle and again 18 hours later. All data are reported as means $\pm$ s.e.m.; *n* $\geq$ 5 for all measurements. For analysis of reproductive function, female mice were checked daily from the time of weaning for vaginal opening and thereafter for vaginal oestrogenization by cellular histology. Median onset of oestrous cycling is reported. Eight-week-old female mice of the genotypes indicated in Table 1 were housed with 8-week-old male *+/+* mice (1 pair per cage). Pregnancy or delivery of pups within 6 weeks was scored as reproductive success and is expressed as fertile animals/animals assayed of each genotype.

Analysis of cell culture and hypothalamic signalling

COS-7 or HEK-293 cells (cultured in DMEM containing 10% FBS) were transfected with pcDNA3Jak2 plus vector control, pcDNA3mLRb, or pcDNA3mLRb^{S1138}, analysed for cell-surface leptin receptor expression, and lysed signalling function essentially as described⁸. For the analysis of hypothalamic leptin signalling, *s/s* mice were calorie-restricted to a similar weight to *+/+* mice by weighing *s/s* and *+/+* mice thrice per week and adjusting the daily food intake of *s/s* mice accordingly. Once weight-matched, all mice were fasted for 48 h to suppress leptin levels further before analysis (leptin levels below 1.0 ng ml⁻¹ were confirmed for animals of all genotypes). Animals were injected intraperitoneally with leptin (100 μ g) or vehicle and were killed 15 min (ERK assays) or 45 min (STAT3 assays) later. Hypothalami were removed and lysed as described⁹. Lysates were resolved directly by SDS–polyacrylamide-gel electrophoresis and immunoblotted with anti-phospho-STAT3 and anti-phospho-ERK; cell lysates were also immunoprecipitated with anti-Jak2 for immunoblotting with anti-PY (ref. 7).

Received 20 November; accepted 10 December 2002; doi:10.1038/nature01388.

- Chua, S. C. Jr *et al.* Phenotypes of mouse *diabetes* and rat *fatty* due to mutations in the OB (leptin) receptor. *Science* **271**, 994–996 (1996).
- Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in *diabetic* mice. *Nature* **379**, 632–635 (1996).
- Chen, H. *et al.* Evidence that the diabetes gene encodes the leptin receptor: Identification of a mutation in the leptin receptor gene in *db/db* mice. *Cell* **84**, 491–495 (1996).
- Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
- Baumann, H. *et al.* The full-length leptin receptor has signaling capabilities of interleukin 6-type cytokine receptors. *Proc. Natl Acad. Sci. USA* **93**, 8374–8378 (1996).
- Stahl, N. *et al.* Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349–1353 (1995).
- White, D. W., Kuropatwinski, K. K., Devos, R., Baumann, H. & Tartaglia, L. A. Leptin receptor (OB-R) signaling. *J. Biol. Chem.* **272**, 4065–4071 (1997).
- Banks, A. S., Davis, S. M., Bates, S. H. & Myers, M. G. Jr Activation of downstream signals by the long form of the leptin receptor. *J. Biol. Chem.* **275**, 14563–14572 (2000).

- Bjorbaek, C. *et al.* Divergent roles of SHP-2 in ERK activation by leptin receptors. *J. Biol. Chem.* **276**, 4747–4755 (2001).
- Elmqvist, J. K., Maratos-Flier, E., Saper, C. B. & Flier, J. S. Unraveling the central nervous system pathways underlying responses to leptin. *Nature Neurosci.* **1**, 445–449 (1998).
- Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
- Chehab, F. F., Lim, M. E. & Lu, R. Correction of the sterility defect in homozygous obese female mice by treatment with the human recombinant leptin. *Nature Genet.* **12**, 318–320 (1996).
- Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early onset obesity in humans. *Nature* **387**, 903–908 (1997).
- Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398–401 (1998).
- Erickson, J. C., Holloper, G. & Palmiter, R. D. Attenuation of the obesity syndrome of ob/ob mice by the loss of neuropeptide Y. *Science* **274**, 1704–1707 (1996).
- Malik, N. M. *et al.* Leptin requirement for conception, implantation, and gestation in the mouse. *Endocrinology* **142**, 5198–5202 (2001).
- Chehab, F. F., Mounzih, K., Lu, R. & Lim, M. E. Early onset of reproductive function in normal female mice treated with leptin. *Science* **275**, 88–90 (1997).
- Elias, C. F. *et al.* Leptin differentially regulates NPY and POMC neurons projecting to the lateral hypothalamic area. *Neuron* **23**, 775–786 (1999).
- Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480–484 (2001).
- Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
- Spaniswick, D., Smith, M. A., Groppi, V. E., Logan, S. D. & Ashford, M. L. Leptin inhibits hypothalamic neurons by activation of ATP-sensitive potassium channels. *Nature* **390**, 521–525 (1997).
- Ollmann, M. M. *et al.* Antagonism of central melanocortin receptors *in vitro* and *in vivo* by agouti-related protein. *Science* **278**, 135–138 (1997).
- Catzellis, C. *et al.* Neuropeptide Y administered chronically into the lateral ventricle profoundly inhibits both the gonadotropic and the somatotrophic axis in intact adult female rats. *Endocrinology* **132**, 224–234 (1993).
- Hahn, T. M., Breininger, J. F., Baskin, D. G. & Schwartz, M. W. Coexpression of Agrp and NPY in fasting-activated hypothalamic neurons. *Nature Neurosci.* **1**, 271–272 (1998).
- Niswender, K. D. *et al.* Intracellular signalling: Key enzyme in leptin-induced anorexia. *Nature* **413**, 794–795 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank G. Leiberman, D. Trombly and M. Seifert for assistance. This work was supported by grants from the NIH (to M.G.M. and M.W.S.), and an ADA/EASD Transatlantic Fellowship (to S.H.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G.M. (e-mail: martin.myers@joslin.harvard.edu).

Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA

Per Arne Aas^{*†}, Marit Otterlei^{*†}, Pål Ø. Falnes^{†‡}, Cathrine B. Vågbo^{*}, Frank Skorpen^{*}, Mansour Akbari^{*}, Ottar Sundheim^{*}, Magnar Bjørås[‡], Geir Slupphaug^{*}, Erling Seeberg[‡] & Hans E. Krokan^{*}

^{*} Institute of Cancer Research and Molecular Biology, Norwegian University of Science and Technology, N-7489 Trondheim, Norway

[‡] Centre of Molecular Biology and Neuroscience, and Institute of Medical Microbiology, University of Oslo, The National Hospital, N-0027 Oslo, Norway

[†] These authors contributed equally to this work

Repair of DNA damage is essential for maintaining genome integrity, and repair deficiencies in mammals are associated with cancer, neurological disease and developmental defects¹. Alkylation damage in DNA is repaired by at least three different mechanisms, including damage reversal by oxidative demethylation of 1-methyladenine and 3-methylcytosine by *Escherichia coli* AlkB^{2,3}. By contrast, little is known about consequences and cellular handling of alkylation damage to RNA⁴. Here we show

that two human AlkB homologues, hABH2 and hABH3, also are oxidative DNA demethylases and that AlkB and hABH3, but not hABH2, also repair RNA. Whereas AlkB and hABH3 prefer single-stranded nucleic acids, hABH2 acts more efficiently on double-stranded DNA. In addition, AlkB and hABH3 expressed in *E. coli* reactivate methylated RNA bacteriophage MS2 *in vivo*, illustrating the biological relevance of this repair activity and establishing RNA repair as a potentially important defence mechanism in living cells. The different catalytic properties and the different subnuclear localization patterns shown by the human homologues indicate that hABH2 and hABH3 have distinct roles in the cellular response to alkylation damage.

Alkylating agents react with DNA at N and O atoms with position-dependent differences and varying chemical reactivity in double-stranded versus single-stranded nucleic acids⁴. The importance of removing alkylation damage from DNA is reflected in the number of repair mechanisms involved. In mammalian cells, the cytotoxic lesion 3-methyladenine is removed by methylpurine DNA glycosylase (ANPG/AAG/MPG), which initiates the multistep base excision repair pathway¹. A different protein repairs highly mutagenic O⁶-methylguanine residues. In this one-step mechanism, the methyl-group is transferred to a cysteine residue in the methylguanine DNA methyltransferase, which is thereby inactivated¹. The oxidative demethylation of 1-methyladenine and 3-methylcytosine in DNA by AlkB is a catalytic mechanism that requires molecular oxygen, Fe²⁺ and 2-oxoglutarate, and the hydroxylated methyl-group is spontaneously released as formaldehyde^{2,3}. The two substrates identified for AlkB are major methylation products in single-stranded DNA and RNA⁴; however, these products are also formed in smaller amounts in double-stranded DNA⁴ and are clearly cytotoxic owing to their capacity to block DNA replication^{2,5,6}.

RNA from prokaryotes and eukaryotes contains numerous naturally occurring methylations and other post-transcriptional modifications. Among 62 naturally occurring RNA modifications known in Eukarya, about 20 are methylations at sequence-specific positions and include 1-methyladenine and 3-methylcytosine (ref. 7 and see <http://medlib.med.utah.edu/RNAmods/>). Enzymatic RNA modifications have been linked to the control of gene expression, in both transcription and translation⁸. They are also important for RNA–RNA and RNA–protein interactions, as well as for the structural stability and correct folding of the RNA⁸. The consequences of RNA methylation damage have not been extensively studied; however, aberrant 1-methylations of adenine in ribosomal RNA in the A-site of ribosomes have deleterious effects on binding of the codon–anticodon complex⁹. In addition, the 1-methyladenine modification results in miscoding by reverse transcriptase¹⁰. Because methylations occurring at the N1 of adenine and the N3 of cytosine modify positions directly involved in base-pairing, they will probably also affect translation. Thus, aberrant methylations in RNA are probably important insults to cellular metabolism and must be actively dealt with by cellular defence systems.

During characterization of the uracil-DNA glycosylase gene (*UNG*)¹¹, we identified an AlkB-like coding sequence immediately upstream of *UNG* (position 12q24.1). We named this gene *hABH2* because a different human homologue (herein called *hABH1*) had been reported previously¹². The small *hABH2* gene spans 5.3 kilobases (kb) and contains four exons. Using *hABH2* as query, we identified a complementary DNA (GenBank accession number NM_139178) and gene for another AlkB homologue that we named *hABH3*. The *hABH3* gene is located at chromosomal position 11q11 close to the centromere and spans 39.4 kb with 10 exons.

The AlkB homologues are relatively small basic proteins with 200–400 amino acids. A sequence alignment shows that the three residues (His 131, Asp 133 and His 187) presumed to constitute the iron-binding cluster in AlkB¹³ are also conserved in the human proteins (Fig. 1a, arrows). On the basis of comparisons with other

well-characterized 2-oxoglutarate-dependent oxygenases, the conserved arginine (Arg 204 in AlkB; Fig. 1a, asterisk) at the start of a conserved block is likely to be involved in binding of C5 carboxylate of 2-oxoglutarate^{13–15}.

We expressed the complementary DNAs encoding hABH2 and hABH3 in *E. coli* and purified the proteins to homogeneity for biochemical studies (Fig. 1b). Biochemical analysis showed that like AlkB, hABH2 and hABH3 removed 1-methyladenine and 3-methylcytosine from DNA by demethylation, as determined by high-performance liquid chromatography (HPLC) of enzyme-treated [³H]methylated DNA (Fig. 2a) and detection of the radioactivity released as formaldehyde (ref. 16 and data not shown). The human enzymes also showed an absolute requirement for Fe²⁺ and 2-oxoglutarate (data not shown). Notably, the substrate specificities for hABH2 and hABH3 were found to be different. As compared with AlkB, which has a moderate preference for single-stranded substrates, hABH2 was much more active on double-stranded DNA (Fig. 2b). By contrast, hABH3 clearly preferred single-stranded substrate DNA.

AlkB and hABH3, but not hABH2, also repaired 1-methyladenine

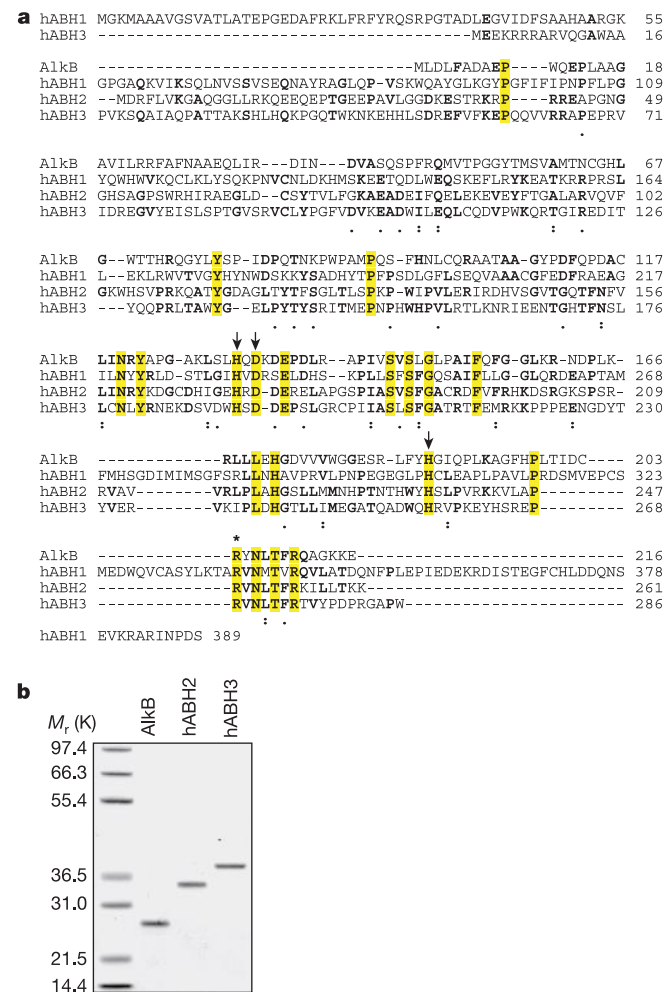


Figure 1 AlkB and human homologues. **a**, Alignment of AlkB, hABH1, hABH2 and hABH3. Identical residues are indicated in yellow, highly conserved residues are indicated by colons, and conserved residues are indicated by dots. Note that a region of the amino- and carboxy-terminal sequences for hABH1 in this alignment is different from the sequence published originally¹². Data are in agreement with all other database entries and our own sequencing results. Arrows indicate residues in the presumed Fe²⁺-binding cluster; an asterisk marks the arginine residue that is likely to be involved in 2-oxoglutarate binding. **b**, SDS-PAGE of purified AlkB, hABH2 and hABH3 proteins.

and 3-methylcytosine in RNA. Repair of alkylated poly(A) by hABH3 was almost as efficient as repair of poly(dA), and repair of poly(C) was substantially more efficient than repair of poly(dA), although not as good as repair of poly(dC) and single-stranded M13 DNA (Fig. 2c). The efficient repair of RNA oligonucleotides strongly indicates that repair of RNA by hABH3 is not a marginal activity and is therefore likely to be biologically relevant. In these experiments, the alkylated substrate contained one lesion in roughly 30,000 nucleotides, which is typical of *in vivo* damage densities¹⁷. Under these conditions, enzyme turnover numbers are low, owing to a large excess of competing non-damaged DNA². Turnover numbers were therefore estimated using as substrate oligonucleotides methylated at high concentration (0.5 M) of methyl methane-sulphonate (MMS). The conversion of 1-methyladenine and 3-methylcytosine to normal bases was determined by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). At a substrate concentration of 0.5–1 μ M, turnover numbers were estimated to be between 0.02 and 0.2 per minute for different substrates.

To examine the effects of AlkB, hABH2 and hABH3 in a cellular

context, we investigated whether *E. coli* DNA and RNA bacteriophages inactivated by MMS could be reactivated by expression of AlkB, hABH2 and hABH3 from a low-copy-number plasmid (pJB658) in the AlkB-deficient *E. coli* host strain HK82 F'. Reactivation of single-stranded DNA phage M13 mediated by hABH2 was similar to that of AlkB, whereas hABH3 expression resulted in partial recovery (Fig. 3a). AlkB and hABH2, but not hABH3, also significantly reactivated methylated double-stranded bacteriophage λ (Fig. 3b). However, the effect of the AlkB-deficiency on host cell reactivation of MMS-treated double-stranded phage was by itself rather small, owing to the much lower yields of 1-methyladenine and 3-methylcytosine versus cytotoxic 3-methyladenine in double-stranded DNA⁴. More significantly, and consistent with the biochemical data, AlkB and hABH3, but not hABH2, also reactivated MMS-treated single-stranded RNA bacteriophage MS2. At the highest dose of MMS, we observed a 10-fold and 100-fold increase in survival of alkylated MS2 after hABH3 and AlkB expression, respectively (Fig. 3c).

We also carried out experiments similar to those described above using purified hABH1 in enzyme assays. Despite having good expression and highly concentrated protein preparations, we failed to show demethylation activity associated with hABH1. In addition, efficient expression of hABH1 in the AlkB-deficient strain used failed to reactivate DNA and RNA bacteriophages (data not shown). Possibly, hABH1 may require processing or modification not taking place in *E. coli* or different *in vitro* assay conditions.

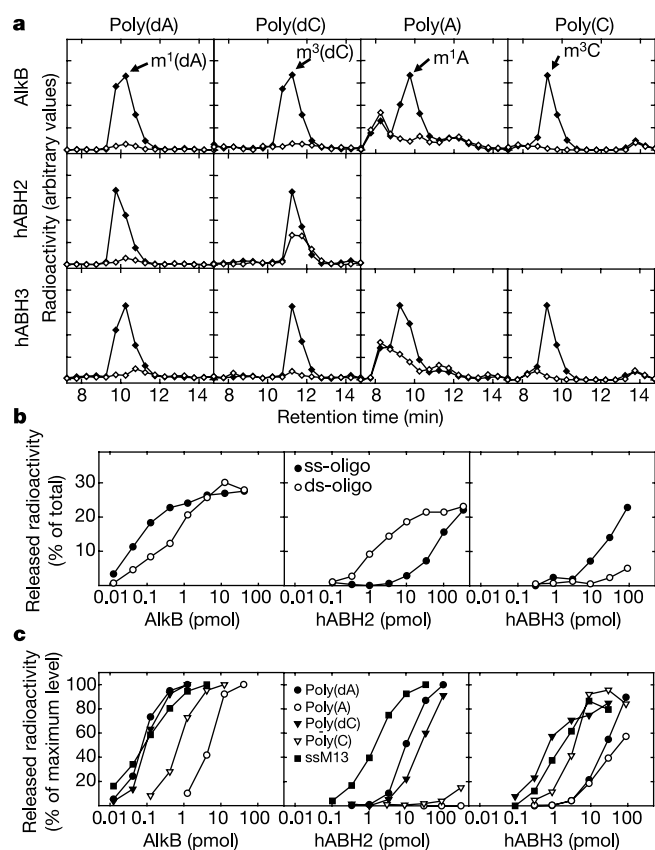


Figure 2 Removal of methylation damage from RNA and DNA by AlkB, hABH2 and hABH3. **a**, Reverse-phase HPLC analysis of [³H]methylated nucleosides in DNA and RNA homopolymers incubated in the absence (filled symbols) or presence (open symbols) of AlkB (top), hABH2 (middle) and hABH3 (bottom). Arrows indicate the identity of the [³H]methylated nucleosides contained in the peaks, on the basis of co-elution with internal standards. **b**, Comparison of activity on single-stranded versus double-stranded DNA. AlkB (left), hABH2 (centre) and hABH3 (right) were incubated with the [³H]methylated, A-rich oligonucleotide 5'-AAAGCAAGAACGAAAAAGCGAAA-3' (filled circles), or with the same oligonucleotide associated with its (unmethylated) complementary strand (open circles), and the released ethanol-soluble radioactivity was measured. **c**, Activity of AlkB (left), hABH2 (centre) and hABH3 (right) on the [³H]methylated substrates M13 single-stranded DNA (filled squares), poly(dA) (filled circles), poly(dC) (filled triangles), poly(A) (open circles) and poly(C) (open triangles). Experiments were carried out at least three times with similar results.

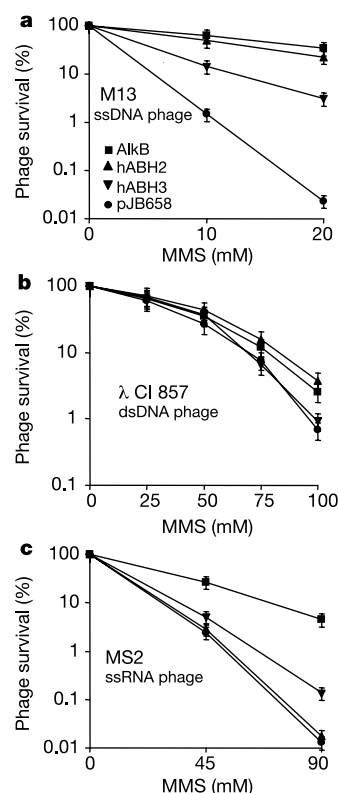


Figure 3 Host cell reactivation of MMS-treated bacteriophages in *E. coli* HK82F' (*alkB*) by episomally expressed AlkB, hABH2 and hABH3. **a**, Reactivation of single-stranded DNA phage M13. **b**, Reactivation of double-stranded DNA phage λ . **c**, Reactivation of single-stranded RNA phage MS2. Squares, AlkB; triangles, hABH2; inverted triangles, hABH3; circles, empty expression vector. Three phage dilutions were plated for each concentration of MMS to ensure adequate plaque numbers. Standard deviations were estimated from 4–6 measurements for each MMS concentration, and the experiments were repeated at least twice with essentially identical results. The number of plaques scored per plate varied from 20 to about 500.

To examine subcellular localization of hABH2 and hABH3 in human cells, we carried out transient transfection experiments using HeLa cells. Both hABH2 and hABH3 seem to be nuclear enzymes (Fig. 4). Outside S phase, hABH2 was homogeneously distributed in the nucleoplasm but showed some accumulation in nucleoli (Fig. 4a). The identity of nucleoli was verified using antibodies against nucleolin (data not shown). In S phase, hABH2 relocates to replication foci, as shown by co-transfection with proliferating-cell nuclear antigen (PCNA), which is known¹⁸ to accumulate in replication foci (Fig. 4a, bottom row). hABH3 was confined mainly to the nucleoplasm, but some staining was observed in the cytoplasm. In most cells, hABH3 was largely excluded from the nucleoli, but occasionally we observed distinct spots in nucleoli (Fig. 4b, top row) and in the nucleoplasm. These spots, however, did not colocalize with PCNA (Fig. 4b, bottom row).

In conclusion, these data indicate that hABH2 repairs DNA close to the replication forks, is more active on double-stranded than on single-stranded DNA and is present in different subnuclear compartments at different stages of the cell cycle. The nuclear localization of hABH3 and the clear preference for single-stranded DNA and RNA suggest that hABH3 has a role in maintenance of nuclear single-stranded DNA and RNA, with genes undergoing transcription potentially being its main targets. The compartmentalization of nuclear processes may demand the presence of more than one AlkB homologue to reduce the detrimental effects of nucleic acid alkylation damage. It seems that the human AlkB homologues hABH2 and hABH3 complement each other such that the sum of their effects matches the activity of AlkB in bacteria.

AlkB homologues have also been identified in plant RNA viruses replicating without DNA intermediates¹³, supporting our observation that RNA repair may be a significant biological function of

AlkB proteins. Damaged RNA is likely to have three fates: it could be ignored, specifically degraded, or repaired. Accurate processing and modification of RNA clearly have a high priority in cells. For example, hundreds of proteins are required to process rRNA correctly¹⁹ and several enzymes are involved in generating the roughly 60 natural modifications of eukaryotic RNA⁷. These modifications are highly important for RNA function and regulation⁸. Therefore, it is not surprising that cells have evolved RNA repair mechanisms that operate to prevent dysfunction of RNA as a result of alkylation damage. In addition, in the extremely crowded macromolecular environment in the cell, about 0.3 g per g cell²⁰, the accumulation of altered proteins resulting from translation of damaged mRNA could result in dominant-negative effects, as well as cytotoxicity from protein aggregates. Our work points to the existence of RNA repair as an important mechanism for cellular defence against such effects. □

Methods

cDNAs for hABH2 and hABH3

We identified the *hABH2* gene located 4 kb upstream of the *UNG* gene¹¹ during sequencing a P1 genome clone. The corresponding cDNA clone (IMAGE ID 46352) was obtained from Research Genetics. cDNA for hABH3 was identified in a BLAST search using cDNA for *hABH2* as query and the clone was purchased (IMAGE ID 4519083) from Research Genetics.

Bacteriophages and *E. coli* strains

The M13 phage stock was prepared as described²¹ using *E. coli* JM109 F' (ref. 22) as the host strain. We produced λ CI857 stock in *E. coli* AB1157. MS2 stock, provided by L. Liljas (Uppsala University), was produced in *E. coli* AB1157 F'.

Purification of AlkB, hABH2 and hABH3 proteins

The *hABH2* and *hABH3* cDNAs and the *AlkB* gene were cloned into the *NdeI/NotI* sites (*hABH2* and *hABH3*) or *NdeI/BamHI* sites (*AlkB*) of the pET28a vector (Novagen) to produce the constructs pET28a-AlkB, pET28a-hABH2 and pET28a-hABH3. His-tagged proteins were expressed in *E. coli* strain BL21(DE3) RIL in 1 l of 3 $\times$ LB medium. Bacterial pellets were sonicated on ice, and the extracts were loaded onto a Ni-NTA Superflow column, which was washed and eluted as recommended (Qiagen). We further purified partially purified proteins by UnoS chromatography (Bio-Rad). Fractions were analysed by SDS-PAGE, and the relevant fractions were pooled. The identity of the homogeneous proteins was verified by Edman sequencing and matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry.

Release of radioactivity by AlkB proteins

N-[³H]methyl-N-nitrosourea-treated DNA or RNA was prepared as described²³. We incubated 50- μ l reaction mixtures³ containing DNA or RNA (typically 0.1 μ g) and enzyme (AlkB, hABH2 or hABH3) for 30 min at 37 °C. DNA or RNA was precipitated by ethanol, and the radioactivity in the supernatant was measured by scintillation counting.

Reverse-phase HPLC analysis of [³H]methylated nucleosides

[³H]methylated DNA or RNA homopolymer (typically 0.5 μ g per 10,000 d.p.m.) was incubated with AlkB (50 pmol), hABH2 (500 pmol) or hABH3 (500 pmol) in a 100- μ l reaction mixture as described above. After incubation, the DNA or RNA was recovered by ethanol precipitation, digested to nucleosides²⁴ and subjected to analysis by reverse-phase HPLC².

LC-MS/MS analyses of alkylated oligonucleotides

We coupled 20 pmol of biotinylated poly(A) and poly(dA) oligomers of 15 nucleotides (MWG-Biotech AG) to magnetic Dynabeads M-280 Streptavidin (DynaL Biotech) and incubated them with 100 μ l of 0.5 M MMS at 30 °C for 30 min. After the MMS was removed, AlkB (0.76–76 pmol), hABH2 (0.6–64 pmol) or hABH3 (0.56–56 pmol) was added to the reaction buffer² and substrate, and the mixture was incubated at 37 °C for 1 h. Nucleosides resulting from nuclease and phosphatase digestion²⁵ were analysed by reverse-phase LC-MS/MS (Sciex API-300, Perkin Elmer) in positive selected ion monitoring mode.

Host cell reactivation of MMS-treated bacteriophages

The *hABH2* and *hABH3* cDNAs and the *AlkB* gene were cloned into the *NdeI/SmaI* (AlkB and hABH2) or the *NdeI/BamHI* (hABH3) sites of the low-copy-number toluic acid inducible vector pJB658 (ref. 26) to produce the constructs pJB658-AlkB, pJB658-hABH2 and pJB658-hABH3. These were transformed into the AlkB-deficient *E. coli* strain HK82F'. M13, MS2 and λ -bacteriophages were treated with MMS essentially as described²⁷. Overnight cultures of transformed strains were diluted in LB medium and induced with 2 mM toluic acid at 37 °C for 3 h. We scored plaque formation in top agar containing the induced strains after infection with different dilutions of MMS-treated phage.

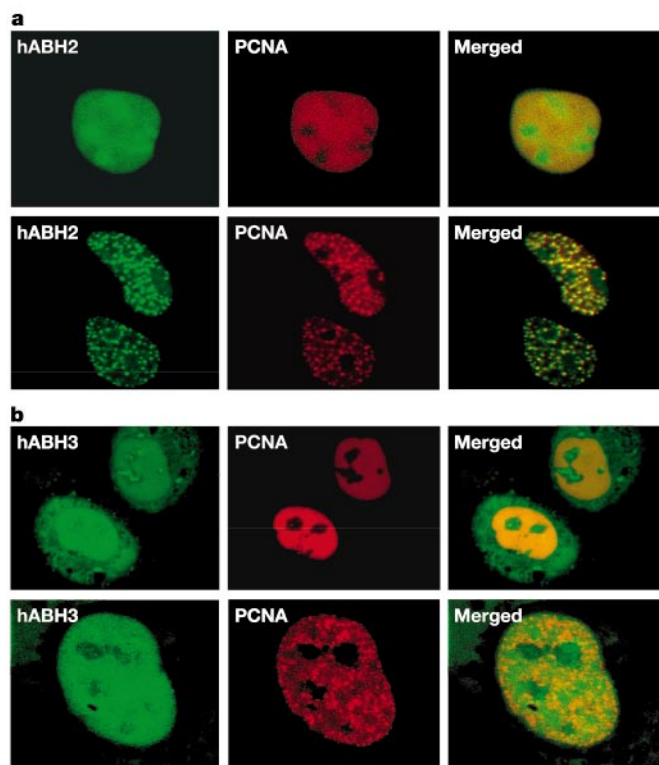


Figure 4 Subcellular localization of hABH2 and hABH3. **a**, Localization of hABH2 (hABH2-EYFP) and PCNA (ECFP-PCNA) in co-transfected cycling living cells outside S phase (top row) and in S phase (bottom row). **b**, Localization of hABH3 (hABH3-EYFP) and PCNA (ECFP-PCNA) in cells outside S phase (top row) and in S phase (bottom row).

Constructs expressing fluorescent fusion proteins

The *hABH2* and *hABH3* cDNAs were cloned, respectively, into the *HindIII/AclI* and *EcoRI/SmaI* sites of pEYFP-N1 (Clontech) to produce hABH2-EYFP and hABH3-EYFP. The pGFP-PCNAL2 construct¹⁸ was used to prepare the expression construct pECFP-PCNA by standard cloning techniques²¹.

Confocal microscopy

Transfection of HeLa cells by the CaPO₄ method was done as described (Profection, Promega). Cells were examined in a Zeiss LSM 510 laser scanning microscope using a 458-nm laser line for excitation of enhanced cyan fluorescent protein (ECFP; detected at 480 nm < λ_{ECFP} < 520 nm) and a 514-nm laser line for enhanced yellow fluorescent protein (EYFP; detected at λ_{EYFP} > 560 nm). Cells were sectioned at a thickness of 1 μm for the images.

Received 8 August; accepted 28 November 2002; doi:10.1038/nature01363.

1. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (ASM, Washington DC, 1995).
2. Falnes, P. O., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**, 178–182 (2002).
3. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).
4. Singer, B. & Grunberger, D. *Molecular Biology of Mutagens and Carcinogens* (Plenum, New York, 1983).
5. Boiteux, S. & Laval, J. Mutagenesis by alkylating agents: coding properties for DNA polymerase of poly(dC) template containing 3-methylcytosine. *Biochimie* **64**, 637–641 (1982).
6. Saffhill, R. Differences in the promutagenic nature of 3-methylcytosine as revealed by DNA and RNA polymerising enzymes. *Carcinogenesis* **5**, 691–693 (1984).
7. Rozenski, J., Crain, P. F. & McCloskey, J. A. The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **27**, 196–197 (1999).
8. Agris, P. F. The importance of being modified: roles of modified nucleosides and Mg²⁺ in RNA structure and function. *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 79–129 (1996).
9. Yoshizawa, S., Fourmy, D. & Puglisi, J. D. Recognition of the codon–anticodon helix by ribosomal RNA. *Science* **285**, 1722–1725 (1999).
10. Matsugi, J. & Murao, K. Study on construction of a cDNA library corresponding to an amino acid-specific tRNA and influence of the modified nucleotide upon nucleotide misincorporations in reverse transcription. *Biochim. Biophys. Acta* **1521**, 81–88 (2001).
11. Haug, T. et al. Human uracil-DNA glycosylase gene: sequence organization, methylation pattern, and mapping to chromosome 12q22–q24.1. *Genomics* **36**, 408–416 (1996).
12. Wei, Y. F., Carter, K. C., Wang, R. P. & Shell, B. K. Molecular cloning and functional analysis of a human cDNA encoding an *Escherichia coli* AlkB homolog, a protein involved in DNA alkylation damage repair. *Nucleic Acids Res.* **24**, 931–937 (1996).
13. Aravind, L. & Koonin, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, Research0007 [online] (<http://genomebiology.com/2001/2/3/research/0007>) (2001).
14. Myllyharju, J. & Kivirikko, K. I. Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J.* **17**, 1173–1180 (1997).
15. Valegård, K. et al. Structure of a cephalosporin synthase. *Nature* **394**, 805–809 (1998).
16. Nash, T. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* **55**, 416–421 (1953).
17. Kaasen, I., Evensen, G. & Seeberg, E. Amplified expression of the *tag*⁺ and *alkA*⁺ genes in *Escherichia coli*: identification of gene products and effects on alkylation resistance. *J. Bacteriol.* **168**, 642–647 (1986).
18. Leonhardt, H. et al. Dynamics of DNA replication factories in living cells. *J. Cell Biol.* **149**, 271–280 (2000).
19. Warner, J. R. Nascent ribosomes. *Cell* **107**, 133–136 (2001).
20. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning. A Laboratory Manual*, 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
22. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
23. Bjelland, S., Björås, M. & Seeberg, E. Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*. *Nucleic Acids Res.* **21**, 2045–2049 (1993).
24. Huang, X., Powell, J., Mooney, L. A., Li, C. & Frenkel, K. Importance of complete DNA digestion in minimizing variability of 8-oxo-dG analyses. *Free Radicals Biol. Med.* **31**, 1341–1351 (2001).
25. Crain, P. F. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.* **193**, 782–790 (1990).
26. Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Haugan, K. & Valla, S. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl. Environ. Microbiol.* **63**, 370–379 (1997).
27. Dinglay, S., Trewick, S. C., Lindahl, T. & Sedgwick, B. Defective processing of methylated single-stranded DNA by *E. coli* AlkB mutants. *Genes Dev.* **14**, 2097–2105 (2000).

Acknowledgements We thank E. Feyzi, K. Baynton and B. Kavli for comments; and A. Nilsen, M. Westbye, R. F. Johansen and L. Hagen for technical support. This work was supported by The Research Council of Norway and the Norwegian Cancer Society. H.E.K. acknowledges support from The Cancer Research Fund at the Regional Hospital in Trondheim, and Svanhild and Arne Must's Fund for Medical Research. E.S. acknowledges support from the European Commission and Simon Fougner Hartmann's Family Fund.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.E.K. (e-mail: hans.krokan@medisin.ntnu.no).

The complete folding pathway of a protein from nanoseconds to microseconds

Ugo Mayor*, Nicholas R. Guydosh*, Christopher M. Johnson*, J. Günter Grossmann†, Satoshi Sato*, Gouri S. Jas‡\$, Stefan M. V. Freund*, Darwin O. V. Alonso||, Valerie Daggett|| & Alan R. Fersht*

* MRC Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, UK

† CLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, UK

‡ Laboratory of Chemical Physics, NIDDK, NIH, Bethesda, Maryland 20892, USA

|| Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195-7610, USA

Combining experimental and simulation data to describe all of the structures and the pathways involved in folding a protein is problematical. Transition states can be mapped experimentally by ϕ values^{1,2}, but the denatured state³ is very difficult to analyse under conditions that favour folding. Also computer simulation at atomic resolution is currently limited to about a microsecond or less. Ultrafast-folding proteins fold and unfold on timescales accessible by both approaches^{4,5}, so here we study the folding pathway of the three-helix bundle protein Engrailed homeodomain⁶. Experimentally, the protein collapses in a microsecond to give an intermediate with much native α -helical secondary structure, which is the major component of the denatured state under conditions that favour folding. A mutant protein shows this state to be compact and contain dynamic, native-like helices with unstructured side chains. In the transition state between this and the native state, the structure of the helices is nearly fully formed and their docking is in progress, approximating to a classical diffusion–collision model. Molecular dynamics simulations give rate constants and structural details highly consistent with experiment, thereby completing the description of folding at atomic resolution.

Engrailed homeodomain (En-HD) from *Drosophila melanogaster* is a 61-residue α -helical protein⁶. Measurement of folding kinetics by temperature jump through the thermal denaturation transition region shows En-HD to be the fastest unfolding and one of the fastest folding proteins so far recorded directly⁴, and its unfolding and folding are compatible with the present time regime of molecular dynamics simulation. Our previous simulations of the unfolding of En-HD predict the correct order of magnitude of the unfolding rate constant and also significant residual native α -helical structuring in the denatured state ensemble⁴. We have now been able to generate a denatured state, by means of mutation, that is stable under physiological conditions, to analyse it in depth and to show by NMR that the wild-type protein populates an equivalent denatured state. We have measured rate constants for folding and unfolding from some hundreds of nanoseconds to microseconds to show that the protein folds by an intermediate that is, in fact, the denatured state under physiological conditions, and we have simulated the unfolding of the protein by molecular dynamics to reconstruct the structures of the transition and denatured states. We now describe each state along the folding pathway, first by experiment and then by simulation.

We first generated a denatured state that was stable under physiological conditions. The native state for most proteins is favoured by only 5–15 kcal mol^{−1}. This balance has been exploited to analyse native and denatured states in equilibrium under

§ Present address: Biosciences Center, Kansas University, Lawrence, Kansas 66047, USA.

folding conditions by NMR⁷. In principle, however, the equilibrium position for any protein can be shifted by mutation so that the denatured state largely predominates under physiological conditions⁸. We lowered the stability of En-HD by mutating a leucine residue to alanine at position 16 (L16A), which deleted stabilizing tertiary interactions without introducing a side chain that was likely to make new specific interactions. Mutation of leucine to alanine tends to lower the stability of proteins by up to 4–5 kcal mol⁻¹ (ref. 9). Wild-type En-HD had a free energy of folding of only 2.5 kcal mol⁻¹. A change in free energy of 4–5 kcal mol⁻¹ therefore should, and did, radically tip the balance so that the protein was predominantly denatured under physiological conditions.

There were no detectable native tertiary interactions in L16A under physiological conditions; see for example, the aromatic region of the ¹H-NMR spectra (Fig. 1). The backbone was very dynamic, but the protein had most of its native α -helical structure, as monitored by circular dichroism (CD) and H α chemical shifts (Fig. 1). The denatured state revealed by L16A had the characteristics of a folding intermediate with much of the correct native

secondary structure. Its radius of gyration increased by 33% compared with the native state (17.0 and 12.8 Å, respectively), as monitored by solution X-ray scattering experiments. The CD and NMR spectra of L16A changed non-cooperatively with increasingly denaturing conditions towards those of a random coil (data not shown).

We then probed the denatured state of wild-type En-HD directly under physiological conditions using native state ¹H/²H-exchange experiments¹⁰. The free energy of what is effectively the complete unfolding of regions of the protein is measured by the rate of exchange of buried backbone amide groups with solvent. The free energy of denaturation of En-HD in ²H₂O measured by conventional procedures for thermal melting was 2.8 ± 0.2 kcal mol⁻¹ at 5 °C. However, the free energy of opening of most of the amide groups in the helices was significantly higher; on average the helices unfolded with a free energy of 3.4 ± 0.1 kcal mol⁻¹, with helix I being the most protected (Fig. 2). There was little protection against exchange for residues 50–55, residues that were highly flexible in L16A (Fig. 1). Thus, the completely open denatured state of En-HD was ~1 kcal mol⁻¹ less stable than the dominant denatured state, which had considerable native α -helical structure and closely resembled the denatured state of the L16A mutant, namely a folding intermediate. Folding intermediates are often the predominant denatured state of a protein under physiological conditions¹¹.

Folding intermediates generated in thermal denaturation simulations, and then quenched to 25 °C, of both the wild-type and L16A proteins had extensive secondary structure and few tertiary contacts (Fig. 3). The overall helical content was 57% for the wild-type intermediate ensemble and 64% for the crystal structure. Break-down of the helix content along the sequence was consistent with the hydrogen protection results: helix I was highly populated with a helix content of 76%, helix II was less stable at 38%, and helix III had a helix content of 51%. Interestingly, however, the intermediate contained two helices much of the time, as the loop between helices I and II adopted helical structure and served to unite the two segments (Fig. 3). Non-native helical structure in the loop between these helices is supported by the NMR chemical shift deviations for L16A (Fig. 1). The radius of gyration of this intermediate state was 30% greater than the native state (14.4 Å for the wild type, 14.2 Å for L16A and 11 Å for the native state), in agreement with experiments on L16A. Further unfolding at high temperature produced a very open unfolded state with little helical structure (21% on average) (Fig. 3). This highly unfolded form had a radius of gyration of 20.0 and 18.8 Å by simulation and experiment (as measured on the acid-denatured state), respectively. Similarly, experiments on En-HD at high temperature or in high concentrations of urea failed to detect regular secondary structure (Fig. 1).

We also detected a folding intermediate from ultra-fast kinetic

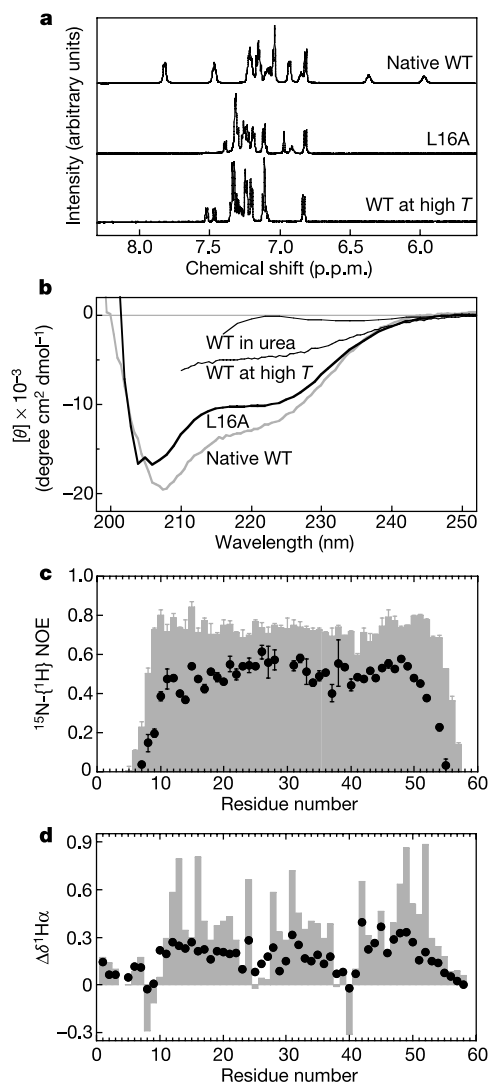


Figure 1 Characterization of wild-type En-HD (WT) and the L16A mutant. **a**, Aromatic region of the ¹H-NMR spectra of WT and L16A. Dispersion of the aromatic chemical shifts was lost in L16A (25 °C). **b**, Far-ultraviolet CD spectra of WT (25 and 95 °C) and L16A (25 °C). **c**, ¹⁵N-¹H heteronuclear NOE data ($\langle r^2 \rangle / r$) of WT (grey bars) and L16A (black circles) at 25 °C. The low NOE values for L16A indicate its higher flexibility. **d**, Deviations in chemical shift from reported random coil values ($\Delta\delta^1\text{H}_\alpha$) in p.p.m. of WT (grey bars) and L16A (black circles) at 25 °C.

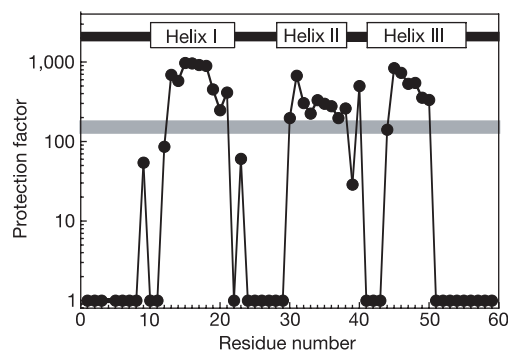


Figure 2 Protection factors for the amide groups of wild-type En-HD, calculated from the ¹H/²H-exchange experiments at 5 °C. A value of 1 is given to those residues for which protection was not observed. The expected protection factor from a two-state fit of the denaturation curves is the thick grey line.

measurements. We showed previously that we could measure the rate constants for folding and unfolding by changing the temperature of a solution of En-HD rapidly (temperature jump) and monitoring the fluorescence of the single tryptophan residue⁴. We

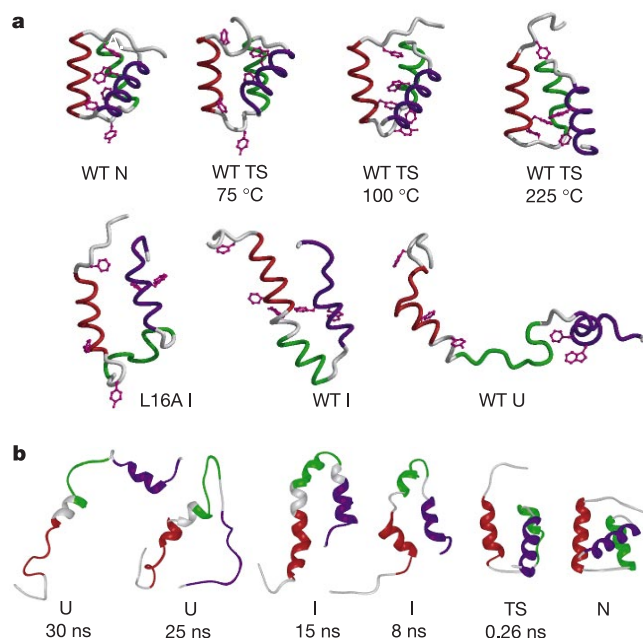


Figure 3 Representative structures from the molecular dynamics simulations. **a**, Snapshots for the transition-state (TS) ensembles identified from wild-type (WT) simulations at different temperatures (61 ns at 75 °C, 1.985 ns at 100 °C and 0.26 ns at 225 °C), the intermediate (I) state (10-ns snapshots from the wild-type and L16A 225 °C trajectories quenched to 25 °C) and the unfolded (U) state (represented by the 40-ns WT 225 °C structure). The native helical segments are coloured as follows: red, residues 10–22; green, residues 28–38; blue, residues 42–55. Aromatic side chains are shown in magenta. **b**, Structures from the wild-type 225 °C denaturation simulation shown in reverse, to illustrate a probable folding pathway of the protein to reach the native (N) state.

have now increased the resolution of our experiments with the use of laser *T*-jump¹² and improvements to the classical electrical discharge apparatus⁴. In so doing, we detected two-step folding kinetics (Fig. 4). At 25 °C, there was a fast phase of $t_{1/2} \approx 1.5 \mu\text{s}$, in addition to the previously reported phase of 15 μs (ref. 4). We were able to assign these phases by two means. First, *T*-jump kinetics on L16A unfolding and refolding exhibited just the fast relaxation time (Fig. 4), indicating that it corresponded to the transition between its equilibrium denatured state (folding intermediate) and a more unfolded denatured state. Indeed, with increasing temperature the tryptophan fluorescence in both the mutant and the native proteins approached that of an exposed tryptophan residue both in equilibrium and *T*-jump experiments. Second, we measured the rate of formation of the wild-type native structure by the change in line shape¹³ of the leucine-16 side-chain resonances between 43 and 56 °C. The rate constants corresponded to those of the slower phase in the *T*-jump kinetics (Fig. 4).

If the fast phase corresponds to the full unfolding of the protein detected by the solvent exchange experiments, then combining the values of the measured equilibrium constants for the overall unfolding and the exchange experiments with the standard formulae for relaxation times gives a minimal kinetic scheme at 25 °C of

$$U \xrightleftharpoons[0.7 \times 10^5 \text{ s}^{-1}]{4 \times 10^5 \text{ s}^{-1}} I \xrightleftharpoons[0.2 \times 10^4 \text{ s}^{-1}]{4 \times 10^4 \text{ s}^{-1}} N$$

where U is the unfolded denatured state, I is the folding intermediate with considerable α -helical secondary structure (or D, the denatured state under physiological conditions) and N is the native state. The results of the various kinetic experiments are compiled in Fig. 4. The unfolding experiments were conducted at up to 65 °C and extrapolated to 100 °C. The times to reach the transition state in the unfolding simulations at 75 and 100 °C fit well with experiment: at 75 °C, the simulated time was approximately 60 ns and the extrapolated $t_{1/2}$ was 330 ns; at 100 °C the times were 2 and 5 ns, respectively (Fig. 4).

We analysed the structure of the transition state between N and I by means of ϕ values (a measure of the degree of native structure in the transition state; 0 implies denatured, 1 implies fully native)

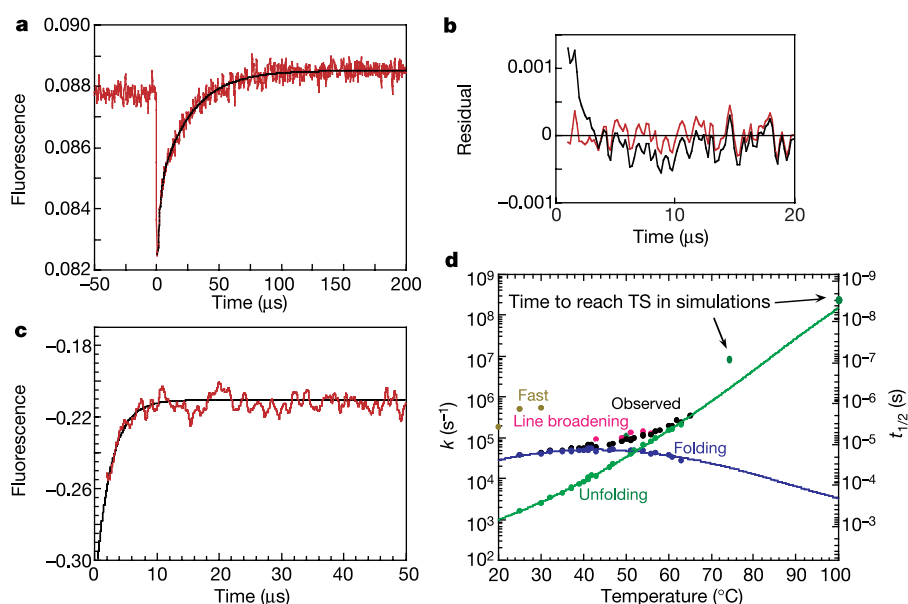


Figure 4 Kinetics of folding and unfolding. **a**, Laser-heating *T*-jump of En-HD to 25 °C; data after 800 ns fitted to a double exponential. Details of output and errors are available from the authors. **b**, Residuals from **a** (red line), randomly distributed around zero, unlike those of a single-exponential fit (black line). **c**, The L16A mutant of En-HD, *T*-jumped to 25 °C (Joule heating), showed only the fast phase. En-HD that had been *T*-jumped with

the Joule-heating apparatus gave very similar rate constants to those from the laser apparatus. Details of output and errors are available from the authors. **d**, Observed relaxation rate constants (k) versus T . Solid curves were derived from combining k with equilibrium data from calorimetry. The filled red circles are results from NMR line-broadening analysis. TS, transition state.

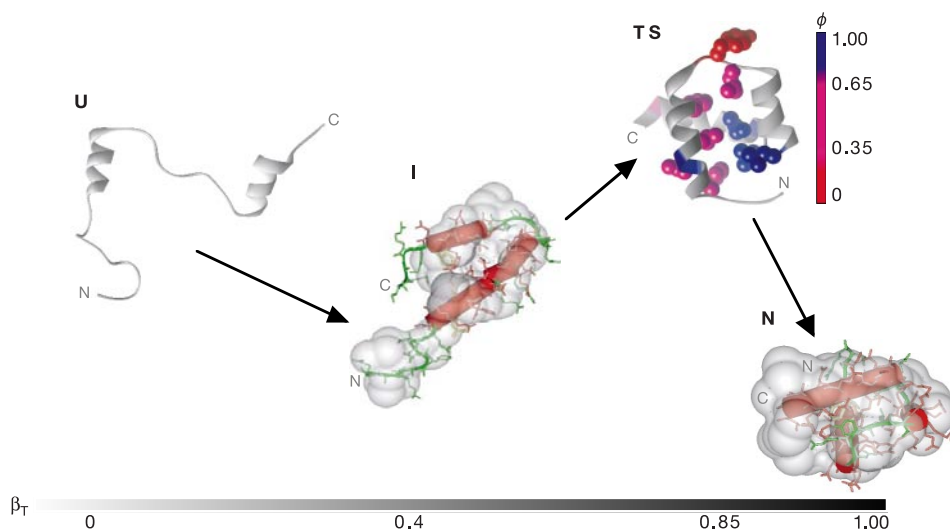


Figure 5 The complete folding pathway of En-HD by experiment and simulation. U is the quenched 40-ns snapshot from the simulation at 225 °C; I is the denatured state under 'physiological conditions'. The 10-ns structure from the high-temperature molecular dynamics simulation is enclosed in the grey molecular envelope obtained by X-ray scattering. TS is the transition state between I and N, the native state. ϕ values are colour-

coded, with residues used as probes of secondary structure coloured directly on the backbone. The simulated structure is from the transition state ensemble for the wild-type at 100 °C (1.985 ns). Structures are positioned according to their relative solvent accessibilities (β_T), as obtained from the dependence of the rate constants on urea concentration. Molecular dynamics simulation gave consistent results.

calculated from the unfolding kinetics¹⁴ for 16 mutations. The transition-state ensembles were also identified from the unfolding simulations at different temperatures, ranging from 75 to 225 °C, and they were very similar (Fig. 3). The experimentally and theoretically derived ϕ values are in good agreement: for example, the correlation coefficient between the values calculated for the 100 °C transition-state ensembles and experiment was 0.86. The combined picture is that the main transition state for folding and unfolding was very native-like, as reflected by both the ϕ values and the molecular-dynamics-generated structures (Fig. 5), as expected for a transition from a structured intermediate. Helix I, helix III and the turn between helices II and III were largely formed (Fig. 3) (for example, ϕ values: A14G in H1 = 0.79; G39A in the turn, 0.92; A43G in H3, 1.05)¹⁴. The loop between helices I and II was less structured (for example, ϕ for A25G was 0.17). The hydrophobic core had fractional ϕ values indicative of partial structure (typically more than 0.4). This structural pattern is just what is expected for a framework¹⁵ or diffusion-collision mechanism¹⁶, in which native-like secondary structure forms and then docks in the rate-determining step (Fig. 3). The use of coarse-grained models has shown that En-HD can fold by diffusion-collision¹⁷. The results here show that a diffusion-collision mechanism is approximated at atomic resolution, both by experiment and simulation by comparing the intermediate and transition states.

The helical contents of isolated synthetic peptides corresponding to helices I, II and III were roughly 30%, 15% and 25%, respectively, as determined both by trifluoroethanol titration experiments¹⁸ and directly from the CD ellipticity at 222 and 208 nm (ref. 19). These abnormally high values, which are probably increased by tertiary contacts in the intact protein, as indicated by our simulations, precipitate a tendency towards the diffusion-collision model. We have argued elsewhere²⁰ that there is a continuum of mechanisms between the classical stepwise diffusion-collision mechanism and nucleation-condensation, in which there is concurrent consolidation of secondary and tertiary structure. It is probably correct to describe En-HD as being close to the diffusion-collision limit, because the pure mechanism of totally independent formation of secondary and tertiary interactions is most unlikely.

En-HD is an informative protein for studying unfolding and

folding. Fast events can be observed directly at the upper limits predicted for folding^{21,22} that enter the time range currently accessible by molecular dynamics simulation. The agreement between molecular dynamics simulation and experiment is very encouraging and on the same timescale. Combining molecular dynamics and experiment has now allowed us to characterize all of the necessary structures along the folding pathway. One of the longest-running controversies in mechanistic studies is whether intermediates are on the pathway or are unwanted side products. Here, simulation shows that the intermediate is on the pathway and that the protein does not fold by an intermediate that must unfold to be productive. Simulation could well be the answer in general for solving such problems in mechanism. □

Methods

Materials

¹⁵N-labelled and ¹³C-¹⁵N-labelled wild-type En-HD were prepared⁴ to >99% homogeneity as determined by both mass spectrometry and SDS chromatography, as was the L16A mutant. CD spectra of 30- μ M samples in 50 mM sodium acetate, pH 5.7, 100 mM NaCl were recorded in an AVIV 202SF instrument. NMR experiments on 0.5–2-mM samples were performed with Bruker DRX 500 and DRX 600 spectrometers equipped with inverse triple-resonance probes and single-axis gradients.

Peptides of sequence AcNH-SSEQLARLKREFNENR-NH₂, AcNH-YLTERRRQQ LSSELGL-NH₂ and AcNH-NEAQIKWIFQNKRAKI-NH₂, corresponding to helices I, II and III of En-HD, were synthesized chemically by G. Bloomberg (University of Bristol). Peptide purity and concentration were determined by mass spectroscopy and amino acid analysis, respectively.

NMR Methods

NMR resonances of both wild-type and L16A mutant En-HD were assigned by using standard triple-resonance techniques²³. ¹H-NMR spectra in ²H₂O, 50 mM sodium acetate-d₃ and 100 mM NaCl were recorded from 1 to 90 °C, with weak presaturation of the residual solvent peak. The spectra were internally referenced to 3-(trimethylsilyl)propionate at 0 p.p.m. Series of ¹H-¹⁵N heteronuclear single-quantum-coherence spectra for solvent exchange experiments into ²H₂O were acquired with 1024 × 90 complex points, four scans per increment and a dead time of 4 min. Solvent exchange data were analysed with standard equations for EX2 conditions²⁴. The apparent equilibrium constant for the formation of the open state was calculated as $K_{op} = k_{ex}/k_{int}$, where k_{ex} is the observed exchange rate and k_{int} is the intrinsic exchange rate calculated from model peptide studies²⁵, with the SPHERE online software²⁶. Apparent free energies of opening were calculated from this equilibrium constant. Enhancements of the ¹⁵N-¹H nuclear Overhauser effect (NOE) were determined from spectra obtained with and without presaturation of amide proton resonances. Presaturation was achieved by applying 120° decoupling pulses in 5-ms intervals with a field strength of 25 kHz for 3 s.

Reference spectra were obtained by replacing the presaturation sequence with a simple delay of the same length. For both experiments, the overall recycle delay was set to 5 s. Two sets of interleaved NOE and control experiments were added to obtain the intensities used in the calculation.

The ^1H -NMR lineshapes of the L16 δ protons from 43 to 56 °C were fitted to a two-site exchange model^{13,27}. Relative populations of the native and the denatured protein were determined independently from differential scanning calorimetry measurements; other parameters were obtained from the spectra. The independent variables were a normalization factor, the exchange rate and the frequency for L16 in the denatured state.

Solution X-ray scattering

Solution X-ray scattering data for wild-type En-HD and the L16A mutant were collected at station 2.1 of the Daresbury Synchrotron Radiation Source, UK, and processed with procedures described previously^{28,29}. Several shape restorations for both protein states have been undertaken to evaluate the similarity and reliability of the low-resolution models. Indeed, the calculations resulted in similar shapes that were consistent with those presented here as semitransparent surfaces.

Kinetics methods

A DIA-RT T-Jump (DiaLog) electrical-discharge temperature-jump instrument¹² and a laser T-jump instrument¹² were used. Protein concentrations were in the range 15–300 μM , buffered with 50 mM HEPES buffer, pH 8.0, 100 mM NaCl. The rate constants were independent of concentration.

Computer simulations

Previous molecular dynamics simulations of the wild-type protein were at 100 and 225 °C for 76 and 42 ns, respectively⁴. New simulations presented here were of the wild type (15 ns at 25 °C, 100 ns at 75 °C, 24 ns at 100 °C, and 21 ns at 225 °C) and the L16A mutant (42 ns at 25 °C, 10 ns at 100 °C, and two independent simulations of 10 and 22 ns at 225 °C). There were further simulations in which structures from these high-temperature simulations were quenched to 25 °C and the water density was adjusted correspondingly (from 0.829 g ml⁻¹ at 225 °C to 0.997 g ml⁻¹). We performed the following quenches: the 10-ns snapshot from the wild-type simulation at 225 °C was simulated at 25 °C for 40 ns to represent the folding intermediate; the 40-ns snapshot from the wild-type simulation at 225 °C was run for 10 ns to represent the unfolded state; and the 10-ns snapshot from the L16A simulation at 225 °C was simulated for 10 ns to model the equilibrium intermediate state. Many of these simulations were performed as controls and to ensure the reproducibility of the results. The various En-HD simulations amounted to 0.422 μs .

Received 18 October 2002; accepted 13 January 2003; doi:10.1038/nature01428.

- Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. Structure–activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**, 6030–6038 (1987).
- Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782 (1992).
- Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27–34 (1996).
- Mayor, U., Johnson, C. M., Daggett, V. & Fersht, A. R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl Acad. Sci. USA* **97**, 13518–13522 (2000).
- Ferguson, N. *et al.* Using flexible loop mimetics to extend phi-value analysis to secondary structure interactions. *Proc. Natl Acad. Sci. USA* **98**, 13008–13013 (2001).
- Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L. & Pabo, C. O. Structural studies of the engrailed homeodomain. *Prot. Sci.* **3**, 1779–1787 (1994).
- Blanco, F. J., Serrano, L. & Forman-Kay, J. D. High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J. Mol. Biol.* **284**, 1153–1164 (1998).
- Spector, S., Rosconi, M. & Raleigh, D. P. Conformational analysis of peptide fragments derived from the peripheral subunit-binding domain from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: evidence for nonrandom structure in the unfolded state. *Biopolymers* **49**, 29–40 (1999).
- Serrano, L. & Fersht, A. R. Principles of protein stability derived from protein engineering methods. *Curr. Opin. Struct. Biol.* **3**, 75–83 (1993).
- Englander, S. W. & Krishna, M. M. G. Hydrogen exchange. *Nature Struct. Biol.* **8**, 741–742 (2001).
- Oliveberg, M. & Fersht, A. R. Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. *Biochemistry* **35**, 2738–2749 (1996).
- Thompson, P. A., Eaton, W. A. & Hofrichter, J. Laser temperature jump study of the helix $\rightleftharpoons$ coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* **36**, 9200–9210 (1997).
- Huang, G. S. & Oas, T. G. Submillisecond folding of monomeric lambda repressor. *Proc. Natl Acad. Sci. USA* **92**, 6878–6882 (1995).
- Matouschek, A., Kellis, J. T. Jr, Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126 (1989).
- Baldwin, R. L. & Rose, G. D. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 77–83 (1999).
- Karplus, M. & Weaver, D. L. Protein-folding dynamics—the diffusion–collision model and experimental data. *Prot. Sci.* **3**, 650–668 (1994).
- Islam, S. A., Karplus, M. & Weaver, D. L. Application of the diffusion–collision model to the folding of three-helix bundle proteins. *J. Mol. Biol.* **318**, 199–215 (2002).
- Kippen, A. D., Arcus, V. L. & Fersht, A. R. Structural studies on peptides corresponding to mutants of the major alpha-helix of barnase. *Biochemistry* **33**, 10013–10021 (1994).
- Greenfield, N. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).
- Daggett, V. & Fersht, A. R. Is there a unifying mechanism of protein folding? *Trends Biochem. Sci.* **28**, 19–26 (2003).
- Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA* **93**, 11615–11617 (1996).
- Bieri, O. *et al.* The speed limit for protein folding measured by triplet–triplet energy transfer. *Proc. Natl Acad. Sci. USA* **96**, 9597–9601 (1999).
- Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego, 1996).
- Itzhaki, L. S., Neira, J. L. & Fersht, A. R. Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J. Mol. Biol.* **270**, 89–98 (1997).
- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
- Zhang, Y.-Z. *Structural Biology and Molecular Biophysics*. PhD thesis, Univ. Pennsylvania (1995).
- Sandstrom, J. *Dynamic NMR Spectroscopy* (Academic, New York, 1982).
- Grossmann, J. G., Hall, J. F., Kanbi, L. D. & Hasnain, S. S. The N-terminal extension of rusticyanin is not responsible for its acid stability. *Biochemistry* **41**, 3613–3619 (2002).
- Witty, M. *et al.* Structure of the periplasmic domain of *Pseudomonas aeruginosa* TolA: evidence for an evolutionary relationship with the TonB transporter protein. *EMBO J.* **21**, 4207–4218 (2002).

Acknowledgements This work was supported with an 'Ikertzaileen prestakuntza' grant (U.M.) from the Government of the Basque Country and with a Winston Churchill Scholarship (N.R.G.). The computational studies were supported by the National Institutes of Health (to V.D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.R.F. (e-mail: arf25@cam.ac.uk).

Screening, scanning and sealing

The kit for drug screening, microplate scanning and microplate sealing.

OmniProbe

Advanced ChemTech

www.advancedchemtech.com

Off-the-shelf peptide libraries

Advanced ChemTech stocks a range of tripeptide, tetrapeptide, protease-targeted peptide and mixture-based penta- and hexapeptide libraries. Omniprobe libraries have been compiled for use as a screening tool in pharmaceutical and biotech laboratories. The libraries include 8,000 individual tripeptides in combinations of all the natural amino acid building blocks and 10,000 individual tetrapeptides comprising most of the natural amino acid building blocks. They are fully compatible with robotic systems.

STN Express 6.01

Chemical Abstracts Service www.cas.org

Patently obvious

The latest version of STN Express 6.01, a program used to search patent, scientific and technical data on the STN international network of over 200 scientific and technical databases, can launch CAS Registry BLAST. This permits users to identify a sequence similar to a queried sequence and view associated patents and published research. It will also carry out STN searches using SciFinder answer keys. Other formatting options, with new pre-defined reports and tables, plus BLAST alerts to help users keep abreast of the latest information, complete the update.

CyBi-SPA Works

CyBio

www.cybio-ag.com

Speed up SPA assays

SPA Works consists of a CyBi well, designed to facilitate liquid handling for pipetting beads, reagents or compounds, and CyBi-Lumas S single-photon counting detection technology. The Lumax 1536 S can detect less than 37.5 µg PVT SPA beads per well in 4 minutes, using clear-bottomed microplates, and has a sensitivity range of 380–650 nm. The system also includes the CyBi PlateSafe for stacking compounds or reagent plates for stand-alone operation.

Spectrum Spotlight

PerkinElmer

www.perkinelmer.com

Pharmaceuticals from the inside out

PerkinElmer has introduced three new products combining FT-NIR (Fourier-transform near-infrared) technology, automated sample handling and powerful software. The Spectrum Spotlight 350 Near-IR imaging system



Keep taking the tablets: Spectrum Spotlight, specifically for drug analysis.

provides a series of pictures of a sample showing chemical distribution at the microscopic level and providing information about micro-scale chemical variations in tablets and solid dosage forms. The Near-IR Tablet Autosampler performs non-destructive whole-tablet assays. Spectrum AssureID 2.0 software simplifies measurement of material quality, verifying the identity and quality of incoming raw materials, intermediates and final products.

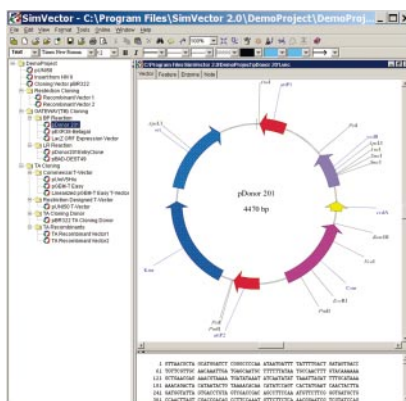
SimVector 2

Premier Biosoft

www.premierbiosoft.com

Cloning simulation and publishing

SimVector 2 aids cloning experimental design and publishing of results. It supports GATEWAY, TA and restriction assay protocols and includes a facility to create custom enzyme sets. Drawing tools are provided to enhance vector maps, which can be exported in vector graphic formats compatible with Adobe Illustrator and PowerPoint. Raster-based images in TIFF, PNG, BMP and web-ready JPEG formats or HTML pages can be generated. SimVector 2 runs on Windows ME/2000/XP and Mac OS X.



Sim Vector2: output for web and print.

High-volume 24-well plates

Porvair Sciences

www.porvair-sciences.com

Automating combinatorial chemistry

These 24-well plates from Porvair Sciences are designed to facilitate rapid, large-scale generation of new target compounds. Using a standard microplate footprint, the plates allow automated, high-throughput target compound preparation in a convenient format. A volume of 10 ml per well allows users to reproduce a large variety of combinatorial chemistries in a single microplate unit.

ProteinScope

Bruker Daltonics

www.bdal.com

Net some storage

ProteinScope is a web-based, client-server platform for storing, organizing and analysing data generated during the entire workflow of Bruker Daltonics' Proteomeer proteomics system. It archives all data generated during the proteomics workflow, from sample separation to protein identification.

Data on separation techniques, such as gel images, spot coordinates and spot characteristics, MS, MS/MS spectra and peak lists, library search results, crosslinks and post-translational modifications are kept in the database. Links to knowledge libraries such as NCI and SWISSProt provide immediate access to additional information.

EZ-2

Genevac www.genevac.co.uk

G-force

Genevac has added a new operating mode to its EZ-2 personal solvent evaporator. The new function allows the unit to operate as a low-speed centrifuge. By disengaging the condenser and pump while allowing the direct-drive motor to run freely, speeds in excess of 2,000 r.p.m. can be achieved. With a standard swing-out rotor, this is equivalent to a force of roughly 500g at the sample — sufficient for many common life-science applications.

TransIT-TKO

Mirus www.genetransfer.com

siRNA transfection reagent in 96-well plates

These plates from Mirus contain siRNA transfection reagent in a 96-well format for screening applications such as target validation. The plates come in four levels, each with a different working concentration of transfection reagent. The reagent is supplied as a dry film in each well of a sterile, clear, flat-bottomed 96-well culture plate. A titration plate contains each of the four concentrations to allow for simultaneous testing as a preliminary screen, since different cell types require different levels of transfection reagent for optimized delivery and knockdown.

FAST PAK 8 & 16 protein array kits

Schleicher & Schuell www.s-and-s.com

Side-by-side on the same slide

These kits contain multiple-pad FAST slides, enabling users to process multiple samples simultaneously on the same slide. FAST PAK kits contain 10 FAST slides, with either 8 or 16 identical pads measuring 6 × 6 mm. Each pad can hold up to 400 elements. The kits work with existing arrayers and scanners, and are compatible with fluorescent, chemiluminescent and colorimetric detection methods.

SealTite microplate seals

TekCel www.tekcel.com

Keep it tight

The SealTite microplate seal, as used on TekCel's PlateServer and PlateStore modules, is designed to enhance automated handling of plates from compound storage through to automated screening systems. SealTite is made of stainless steel with an inner Teflon

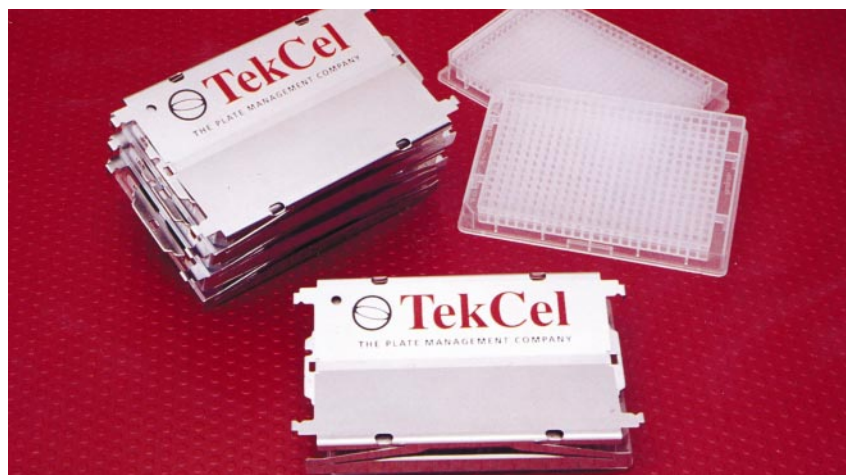


Plate management: TekCel's new sealing technology.

liner compatible with DMSO, and suitable for 96- and 384-well microplates. The seal is removed vertically to minimize well-to-well cross-contamination, and to protect the user from exposure to unknown chemicals.

Heat transfer inserts

Genevac www.genevac.co.uk

Warm your plates

Although centrifugal evaporation is the method of choice for drying delicate samples in drug discovery chemistry, removing solvent in the microplate format can be time-consuming as little of the surface area of the plate is in good thermal contact with the sample holder. Genevac's new heat transfer plates are tight-fitting inserts beneath the plastic microplate that facilitate efficient thermal energy transfer from the hot outer surface of the sample holder directly to the walls of the microplate. By specifying an evaporator with high-power lamps, overall drying times for aqueous samples can be reduced by four-fifths.

Glycoprotein staining kit

Amresco www.amresco-inc.com

Sensitive detection

Amresco's glycoprotein staining kit has been developed for the company's Pro-Pure proteomics line for the detection and staining of glycoproteins on SDS-PAGE gels. The three-component kit is sufficient to stain eight gels, providing a low background for each. The kit has a quick protocol and is convenient and easy to use.

SuperScript III RT

Invitrogen www.invitrogen.com

Be more specific

SuperScript III RT has a half-life of 220 minutes to produce higher yields of cDNA for applications such as array labelling and library construction. It is fully active at 50 °C

for increased specificity when performing RT-PCR with a gene-specific primer. In addition, increasing units in the RT reaction does not inhibit subsequent PCR, a feature that is useful in real-time applications.

These notes are compiled in the Nature office from information provided by the manufacturers. Press releases to be considered for publication can be sent to newproducts@nature.com.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Sille Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakakamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Capturing competencies

Discerning how many discrete skills are involved in drug discovery is a bit like trying to guess how many licks it takes to finish a lollipop. But this month the Science, Technology and Mathematics Council, a nonprofit organization that advises the UK government on science policy, began a project to determine the former, with the kind of resolve required to complete the latter.

The council is surveying 35 of Britain's drug-discovery firms to see what tasks each member of their drug-discovery teams performs, decide what skills are involved, and then set standards for each task. The hope is that the survey, which is likely to take several years, can guide companies on how to judge the competencies of employees working in different parts of the drug-discovery pipeline. It could also help universities to tailor their curricula to industry's needs.

The challenge is that, unlike a lollipop, the drug-discovery pipeline is a moving target. Too much focus on specific skills could make the surveyors lose track of the bigger picture. "What we don't want to do is develop something that will be redundant in six months," says Richard Smith, the council's director of science and technology.

But recommending competencies doesn't necessarily mean that they will be adapted or adhered to. The council conducted a similar exercise in 1995 for forensics, which is now increasingly attractive to life scientists who want to apply DNA technology to crime-solving (see page 872). The resulting competencies were taken seriously by some programmes, says Smith, but others merely paid lip service to them. Nevertheless, discerning students soon learned which courses delivered the skills they needed. Hopefully, the council's latest survey will similarly empower students if they choose to take an industrial path.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

A trip to the crime labs p872

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Dianne Burns finds forensic science to be an interesting use for her science qualifications.

Forensics offers career opportunities for the Sherlock Holmes of the future – and the work is far from elementary, says Susan Myers.

Geneticist Patrick O'Donnell was happy to help police with their enquiries when they rang him in 1990. During a postdoc at San Diego State University, O'Donnell had started to question whether he wanted to remain in academic research. The San Diego Police Department (SDPD) Crime Laboratory was looking for a scientist to coordinate the design and construction of a forensic DNA-analysis laboratory. It was a perfect match.

O'Donnell's project, studying defects in the muscles of fruitflies, was "so esoteric and uninteresting to the general population", he says, "that I found myself asking: 'What contribution am I making?'" But work in the SDPD Crime Laboratory offered him both a chance to 'contribute' and a degree of job security. Although he had some trepidation about 'selling out' of academic research and having to build a new facility and scientific programme from scratch, he applied for the post and got the job. He doesn't regret the move. "Everything my lab does every day touches people," O'Donnell says.

Highly trained scientific specialists such as O'Donnell are relatively new to crime labs, which have traditionally been home to generalists who process evidence such as hair, fibres, soil and fingerprints. But advances in the technologies behind DNA analysis — particularly in the analysis of individual genetic variation found at short tandem repeats (STRs) in the genome — have enhanced a researcher's ability

to identify a suspect by matching DNA evidence from a crime scene to a reference sample. And the increase in the number of DNA databases containing such evidence has made the technique all the more powerful.

This new brand of DNA analysis, and the attendant databases that have sprung up around it, has opened up new career possibilities for scientists. The number of forensics labs already in existence hint at the opportunities — the United States has 120 public forensic laboratories nationwide, and the United Kingdom has seven regional labs. California has been especially active in seeking out specialists. The DNA lab at California's Department of Justice (DOJ) has played a large part in training forensic DNA analysts. In a push to populate its database in 2001, the DOJ hired about 100 criminalists to analyse hundreds of DNA samples a day.

BACK TO THE LAB

One beneficiary was Dianne Burns, who says she had always been interested in science. But because of "lack of discipline", she dropped out of college to become a carpenter's apprentice in British Columbia in the 1980s, moving to California to build houses before realizing she needed more intellectual stimulation.

Burns finished her biology degree at the University of California, Davis, then worked for a year in the coroner's office before taking a job as a technician identifying and locating fingerprints in the San Mateo County Sheriff's Department. In December 1991, she became a criminalist with the DOJ DNA lab in Berkeley. She has worked there ever since — apart from a one-year sabbatical to complete a masters degree in forensic science at King's College London, during which she spent some time working for the police in Edinburgh, where she investigated the chemistry of arson accelerants. Burns now says that if she had the chance to do her degrees over again, she might opt to study chemistry, but "you have to go where the money is", and right now the forensic cash is in DNA analysis.

Heightened public awareness of the significance of DNA evidence in criminal prosecutions has also affected the funding decisions made by legislators. California recently received \$46 million for the California Cold Hit programme, which uses STR analysis on evidence in suspectless cases and in unsolved sexual assaults and murders.

Forensic-education programmes are now cropping up to accommodate the increase in public interest, says

Expert witness

Robert Shaler, director of forensic biology at the Office of the Chief Medical Examiner (OCME) in New York, eased into the world of forensic science as an expert witness.

Shaler had completed his PhD in biochemistry, worked as a postdoc at the University of Pittsburgh School of Medicine, and then taught at the School of Pharmacy at the University of Pittsburgh. Along the way, and "just out of interest", he took night courses in forensic science.

He eventually took a job at the Pittsburgh crime lab, where he

received a research grant to work on a private case from New York City, for which he was asked to testify.

Shaler's work on the case was reviewed by the OCME, where it was so highly regarded that he was later offered a job when a position became available.

Shaler now employs 105 people, in positions ranging from assistant director to technician. When hiring staff, Shaler opts for a careful balance. "I don't want a homogeneous lab," he says. "I shy away from taking people all from the same background or programme."

S.M.



Following procedure: (from right) Patrick O'Donnell, Amy Rogala and intern Vonda McAllister assess the evidence in San Diego.

Burns. But she advises prospective students to choose an established, reputable programme that teaches a set of competencies endorsed by the government, rather than a more general course. In Britain, two stood out for her: one at the University of Strathclyde in Glasgow and the one she took at King's College.

Barbara Daniel, director of the King's College degree programme, says that the large number of applicants for the course allows her to be choosy. She prefers students who have degrees in molecular biology or chemistry, with some coursework in statistics and laboratory experience. Many are on study leave from their police or forensic jobs. The King's College programme admits 40 students a year out of 200 or so applicants, with about a quarter of the students coming from abroad.

DETERMINED APPROACH

Not all scientists are cut out for forensics, which requires strict attention to detail. Daniel calls it "dogged determination", and says that successful criminalists are not necessarily those who understand all of the theory, but the ones who can communicate science to the public. "You have to be a good communicator to be able to stand up in court," she says. The innovative nature of many research scientists would be a disadvantage, as criminalists must follow standard operating procedures without variance.

Ian Fitch, who swapped from a postdoc at the University of California, Santa Barbara, to the SDPD forensic-biology unit, agrees. "PhD training requires that you 'think outside the box,'" he says, "whereas forensic science is often 'inside the box.'" As a rule, the profession attracts people with certain characteristics: criminalists are a meticulous lot.

"Everyone in the field is compulsive," says Amy Rogala, a former biology teacher who switched to a masters degree in forensics before she was hired as a criminalist by the California DOJ. Joking that she

studies human behaviour, she tells of colleagues repeatedly straightening papers on their desks, and of envelopes discarded because their ink labels had faintly transferred the words 'freezer packet' from the front of one to the back of another. Rogala includes herself in this conscientious group. "I'm a finisher," she says. "I started a case yesterday, and I finished it yesterday."

Fitch agrees. He likes the fact that there is a beginning, a middle and an end to the work. "In research," he says, "there's a beginning, and a beginning..." For O'Donnell, the satisfaction comes from realizing that the work done by the criminalists he has hired contributes daily to the lives of people and changes the outcome of criminal cases. "It was the best decision I ever made," he says. "I never go home any more and say 'what am I doing with my life?'" ■

Susan Myers is a scientific adviser to Cozen O'Connor Attorneys in San Diego.

The future of forensics

Following protocols and standard operating procedures is crucial to forensic science, but technological innovations are still improving the tools of the trade. In some cases, the DNA evidence is too scant for the common analytical methods. In these instances, the FBI and the Office of the Chief Medical Examiner use mitochondrial DNA analysis, which examines maternally inherited DNA found in specific parts of the cell. Although it is expensive, other labs may soon adopt this procedure.

Another emerging technology

uses a larger amount of DNA in the analysis of single nucleotide polymorphisms (SNPs) — points in the genome at which the sequence varies between individuals by a single letter in the genetic code. A collection of SNPs from a sample could shift DNA analysis from determining a suspect's identity towards the predictive realm. For example, in a case with no suspects and no match in the DNA databases, SNPs from the sample could provide clues to physical characteristics and ethnicity, offering a genetic profile that could be used to track down a suspect. S.M.